BIOESTATÍSTICA

SONIA VIEIRA

TÓPICOS AVANÇADOS

Testes não paramétricos,
Testes diagnósticos,
Medidas de Associação e
Concordância.

4ª EDIÇÃO

ELSEVIER

BIOESTATÍSTICA

TÓPICOS AVANÇADOS

4ª EDIÇÃO

Sonia Vieira
Doutora em Estatística pela USP
Livre-docente em Bioestatística pela Unicamp
Pós-doutorado em Estatística na Universidade da Califórnia, Berkeley, Califórnia
Pós-doutorado em Estatística na Universidade Yale, New Haven, Connecticut
Pós-doutorado em Ética na Schloss Leopoldskron, Innsbruck, Áustria

ELSEVIER

ISBN: 978-85-352-8981-7
ISBN versão eletrônica: 978-85-352-8982-4

Capa: Monika Mayer e Luciana Mello
Editoração Eletrônica: Rosane Guedes.
Epub: SBNigri Artes e Textos Ltda.

Elsevier Editora Ltda.
Conhecimento sem Fronteiras
Rua Sete de Setembro, 111 – 16º andar
20050-006 – Centro – Rio de Janeiro – RJ – Brasil

Rua Quintana, 753 – 8º andar
04569-011 – Brooklin – São Paulo – SP – Brasil

Serviço de Atendimento ao Cliente
0800-0265340
atendimento1@elsevier.com

Consulte nosso catálogo completo, os últimos lançamentos e os serviços exclusivos no site www.elsevier.com.br

NOTA
Esta obra foi produzida por Elsevier Brasil Ltda. sob sua exclusiva responsabilidade. Médicos e pesquisadores devem sempre fundamentar-se em sua experiência e no próprio conhecimento para avaliar e empregar quaisquer informações, métodos, substâncias ou experimentos descritos nesta publicação. Devido ao rápido avanço nas ciências médicas, particularmente, os diagnósticos e a posologia de medicamentos precisam ser verificados de maneira independente. Para todos os efeitos legais, a Editora, os autores, os editores ou colaboradores relacionados a esta obra não assumem responsabilidade por qualquer dano/ou prejuízo causado a pessoas ou propriedades envolvendo responsabilidade pelo produto, negligência ou outros, ou advindos de qualquer uso ou aplicação de quaisquer métodos, produtos, instruções ou ideias contidos no conteúdo aqui publicado.

CIP-BRASIL. CATALOGAÇÃO-NA-FONTE
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ

Vieira, Sonia, 1942-
Bioestatística : tópicos avançados / Sonia Vieira. - 4. ed. - Rio de Janeiro : Elsevier, 2018.
il. ; 24 cm.

V718b
4. ed.

Inclui bibliografia
ISBN 9788535289817

1. Bioestatística. I. Título.

18-50011

CDD: 570.15195
CDU: 57.087.1

Meri Gleice Rodrigues de Souza - Bibliotecária CRB-7/6439
24/05/2018 01/06/2018

Escribo como un loco para no volverme loco.
Leonardo Padura

Prefácio

Com a popularização dos computadores, as estatísticas passaram a fazer parte do vocabulário da área da saúde. Quem lê artigos nessa área já ouviu falar em testes não paramétricos, fator de risco, razão de chances e testes diagnósticos. Não basta, porém, que o profissional de saúde tenha "ouvido falar" de Bioestatística; é preciso que adquira visão adequada sobre o assunto, entendendo, por exemplo, o que um programa de computador pode fazer por ele – sem expectativas excessivas.

Este livro é, basicamente, uma continuação do livro *Introdução à Bioestatística*. Por essa razão, tanto pode ser útil para quem está estudando Estatística como pode ser útil para profissionais que atuem como consultores nas diferentes áreas de saúde. A falta de literatura em português sobre alguns dos temas tratados aqui também pode ser uma motivação para a leitura.

Mas escrever um livro não é tarefa fácil. Precisei da ajuda de pessoas diferentes, que me ajudaram de maneiras diferentes. Quero, portanto, agradecer à minha amiga Martha Maria Mischan, que muito sabe sobre Estatística. Ela leu os manuscritos e apontou erros, com maestria e cordialidade. Outras pessoas também me ajudaram: William Saad Hossne fez a apresentação da obra com a disposição de quem sempre esteve pronto a me ajudar, José Eduardo Corrente respondeu às minhas muitas perguntas e Gláucia Melo gentilmente acertou as minhas referências. Também devo agradecimentos a meu filho, Márcio Vieira Hoffmann, que sempre me ajudou.

Mas meus maiores agradecimentos são para todo o pessoal da Elsevier Editora, pelo apoio constante ao meu trabalho. Agradeço, ainda, aos muitos professores e colegas que propiciaram o ambiente para que eu pudesse escrever. No entanto, um professor e escritor não existe sem seus alunos e seus leitores. Devo, portanto, enorme agradecimento a quem me lê.

A autora

Sumário

1

Dados, Variáveis e Outros Termos

Para muitas pessoas, a palavra Estatística lembra números. No sentido de informação numérica, as estatísticas já fazem parte de nosso dia a dia. Discutimos – com base em informações numéricas – o aumento da expectativa de vida, a diminuição das taxas de natalidade ou a preocupação das seguradoras de saúde com o aumento de casos de obesidade. E lemos, em jornais e revistas, que o estresse é fator de risco para o infarto do miocárdio ou que a terapia com células-tronco tem grande probabilidade de sucesso.

Estatística não é, porém, simples coleção de números, embora as medidas ou observações na forma numérica sejam sua base. No sentido acadêmico, a Estatística pode ser dividida em duas partes:

- *Estatística Descritiva*, que são os procedimentos utilizados para organizar a coleta, a apuração, a classificação e a descrição dos dados.
- *Inferência Estatística*, que são os procedimentos através dos quais se generaliza o conhecimento adquirido em amostras para a população de onde as amostras provieram.

1.1. DADOS E VARIÁVEIS

Dado é a informação coletada e registrada, referente a uma variável.

Variável é uma condição ou característica que descreve uma pessoa, um animal, um lugar, um objeto, uma ideia. A variável pode assumir valores diferentes em diferentes unidades.

As variáveis – e consequentemente os dados – são classificadas em dois tipos:

- Categorizadas (antes chamadas qualitativas).
- Numéricas (antes chamadas quantitativas).

Variável categorizada ou qualitativa: seus valores são expressos em palavras.

Variável numérica ou quantitativa: seus valores são obtidos por medição ou contagem.

Exemplo 1.1

Raça de cães é variável categorizada (pequinês, pastor alemão, fila, boxer etc..)

População de uma cidade é variável numérica.

As variáveis categorizadas ou qualitativas são classificadas em dois tipos:

- Nominal.
- Ordinal.

Variável nominal: seus valores se distribuem em categorias mutuamente exclusivas, indicadas em *qualquer* ordem.

Variável ordinal: seus valores se distribuem em categorias mutuamente exclusivas que têm ordem intrínseca.

Exemplo 1.2

O resultado de um teste de gravidez é variável nominal porque pode ser "positivo" ou "negativo", mas as possibilidades de resultado também podem ser verbalizadas em outra ordem: "negativo" ou "positivo".

O grau de satisfação de um paciente com a atenção recebida do profissional da saúde é variável ordinal: "muito bom", "bom", "regular", "ruim", "péssimo". A ordem pode ser invertida ("péssimo", "ruim", "regular", "bom", "muito bom"), mas não é colocada em *qualquer* ordem.

As variáveis quantitativas ou numéricas são classificadas em dois tipos:

- Discreta.
- Contínua.

Variável discreta: só pode assumir alguns valores em um dado intervalo.

Variável contínua: pode assumir qualquer valor em um dado intervalo.

Exemplo 1.3

São variáveis discretas: número de filhos, número de batimentos cardíacos por minuto, número de dentes presentes na boca.

São variáveis contínuas: peso, tempo de espera numa fila, temperatura corporal.

A Figura 1.1 resume a classificação das variáveis.

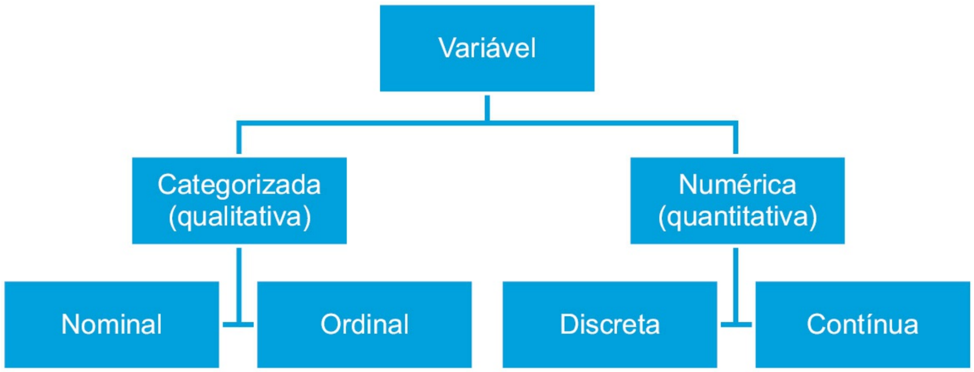


Figura 1.1 – Tipos de variáveis.

Dados referem-se às variáveis, mas são valores coletados por observação, medição, questionamento. Cada unidade investigada fornece seus dados. Coletamos dados como, por exemplo, a idade de brasileiros, para conhecer a distribuição dessa variável.

Exemplo 1.4

As informações contidas em cédulas de identidade – como nome completo do cidadão, data e local de nascimento – são variáveis, mas em *sua* cédula de identidade estão *seus dados*.

1.2. CUIDADOS NO REGISTRO DOS DADOS

1.2.1. Dados nominais e ordinais

Dados nominais e ordinais são descritos por palavras, mas podem ser registrados por números. Esses números não têm, no entanto, significado numérico. Por exemplo, muitas vezes sexo é registrado como 1 para masculino e 2 para feminino. Isso não significa ordem nem valor numérico (sexo feminino não é "segundo sexo", nem uma mulher vale o dobro de um homem).

Em algumas áreas, é usual, ou até mesmo convencional, registrar dados ordinais por números.

Exemplo 1.5

Os tumores são estadiados de acordo com o grau de desenvolvimento em uma escala numérica, que em geral varia de 0 a IV. Os números indicam apenas ordem: a Fase IV vem depois da Fase II porque tem pior prognóstico, mas a Fase IV não significa o dobro da Fase II.

Quando as categorias de uma variável ordinal são referenciadas por números, é preciso que a pessoa que faz o registro do dado tenha definição clara do significado de cada número que compõe a escala de medida da variável.

Exemplo 1.6

A Organização Mundial da Saúde (OMS) desenvolveu o Sistema de Medição da Autonomia Funcional para avaliar a atividade funcional das pessoas. A pessoa entrevistada responde a um questionário com 29 perguntas. O entrevistador precisa saber o significado de cada número:

- 0 = independente;
- 1 = precisa de supervisão;
- 2 = precisa de ajuda;
- 3 = dependente.

A soma dos pontos nas 29 perguntas é uma medida da atividade funcional da pessoa.

Na maioria das vezes, porém, as categorias das variáveis ordinais não são definidas de maneira única e convencional. É o pesquisador quem define a escala para estabelecer as categorias da variável que estuda. Por exemplo, pode ser definida uma escala de três pontos para grau de satisfação: "bom", "regular", "ruim". Em lugar de palavras, o pesquisador também pode usar figuras, notas ou uma escala visual analógica.

Figuras: são usadas para obter respostas de pessoas com pouca escolaridade ou de crianças, mas – para relatar e avaliar os resultados da pesquisa – os pesquisadores transformam as figuras em palavras.

Exemplo 1.7

Pode-se perguntar a uma criança: “Que cara você faz quando vai ao dentista?” e pedir que ela aponte uma das figuras abaixo:

Nota: em centros de atendimento, é comum solicitar ao participante de pesquisa que atribua uma nota ao que está sentindo, seja dor, medo, desconforto ou, até mesmo, uma nota à satisfação com um procedimento.

Exemplo 1.8

Pode-se perguntar a um paciente que se submeteu a uma cirurgia estética: “Em uma escala de 0 a 5, que nota você daria ao resultado da sua cirurgia?”

Escala visual analógica (EVA): para usar a escala visual analógica, o pesquisador fornece ao participante de pesquisa o desenho de um segmento de reta de comprimento conhecido. No início e no final desse segmento devem estar escritos estados extremos do sentimento que o pesquisador pretende medir. Por exemplo, devem estar escritos “péssimo” no início do segmento e “excelente” no final. O participante da pesquisa marca um “X” na alternativa que mostra o que está sentindo. Depois, o pesquisador mede a distância do início do segmento até o “X” – e tem um número.

Exemplo 1.9

Para avaliar a dor após instrumentação cirúrgica, o dentista pode solicitar a seu paciente que indique o grau de dor que está sentindo em uma escala como a apresentada em seguida:

Nenhuma dor ______________________ Dor insuportável

Quando se usam figuras, notas e escalas visuais, cada pessoa avalia a si própria. Então, os dados para análise são valores conferidos por pessoas diferentes avaliando situações diferentes. A nota 3 para uma pessoa pode *não* ter o significado que a mesma nota 3 tem para outra pessoa. É preciso cuidado na análise.

1.2.2. Dados discretos e dados contínuos

Dados discretos *não podem ser confundidos* com dados ordinais. Por exemplo, se alguém apresenta uma distribuição de mulheres segundo o número de filhos nascidos vivos (0, 1, 2, 3, 4 ou mais), está relatando um dado discreto (número de filhos são contáveis), mas se alguém apresenta uma distribuição de mulheres com câncer de mama segundo o estadiamento do tumor

(Fases 0, I, II, III e IV), está relatando um dado ordinal. Pode ser calculado o número médio de filhos por mulher, mas para a análise da distribuição de estadiamento clínico de tumores malignos deve ser construída uma distribuição de frequências.

Os valores que podem ser assumidos por uma variável contínua são limitados pela precisão do instrumento de medida.[1] Algumas vezes, porém, é usual registrar dados com poucos decimais. Por exemplo, nos estudos com adultos registram-se idades em anos completos, embora a idade seja uma variável contínua (envelhecer é um processo contínuo). Somente nos estudos com pré-escolares registra-se a idade em meses completos e somente nos estudos com recém-nascidos registra-se a idade em dias.

De qualquer forma, os dados numéricos devem ser registrados na forma em que foram obtidos – e não por classes ou categorias. Registre a idade da pessoa – e não seu grupo de idade. Registre o peso, a pressão sanguínea, o nível de colesterol do paciente como foram medidos e só depois, se for o caso, converta em classes para análise. Não há como recuperar dados contínuos que foram registrados em classes. No entanto, como toda regra, esta também tem exceção: se a variável coletada é altamente imprecisa, como número de cigarros fumados por dia, é mais sensato anotar os dados apenas por categorias, como: nenhum, de 1 a 5, de 6 a 10, de 10 a 20, 21 ou mais.

1.3. DADOS DISCREPANTES, PERDIDOS E CENSURADOS

Os artigos científicos por vezes fazem referência a dados discrepantes (*outliers*), dados perdidos (*missing values*) ou dados censurados (*censored data*). Nesta seção, é feita breve referência a esses tipos de dados.

Dado discrepante ou **atípico** é um valor extremo, que difere muito dos demais dados da amostra (por ser muito maior ou muito menor do que os outros).

É importante saber que dados discrepantes têm influência considerável nos resultados dos testes paramétricos. A inclusão – ou a exclusão – de tais dados tem efeito real sobre a média e a variância e pode, por conta disso, modificar o resultado de um teste estatístico. Uma estratégia razoável para enfrentar a situação é analisar os dados incluindo, depois excluindo, os dados discrepantes. Se os resultados forem praticamente iguais, use a análise que inclui os dados discrepantes. Se os resultados mudarem muito com a inclusão dos dados discrepantes, procure uma *explicação* – que possa justificar a inclusão ou a exclusão desses dados na amostra. Outra solução – proposta neste livro – é o uso de um teste não paramétrico.

De qualquer forma, saiba que alguns analistas insistem na ideia de que todo valor que se distancie mais de três desvios padrão da média deve ser descartado. Outros calculam a distância interquartílica e consideram que deve ser descartado todo valor que esteja 1,5 vez a distância interquartílica acima do terceiro quartil ou abaixo do primeiro quartil. Embora as técnicas estatísticas possam ser usadas para apontar os dados discrepantes, elas *não* devem ser usadas para determinar *o que fazer* com eles.

De qualquer modo, quando o pesquisador se depara com um dado discrepante, deve verificar se houve erro no registro ou na coleta desse dado. Se ficar comprovado que o dado está errado porque houve falha em seu registro ou na coleta, deve ser descartado. Usando essa mesma lógica – de que falhas no processo de registro ou de coleta justificam o descarte –, também deve ser retirado da análise todo dado que, embora plausível, não esteja de acordo com critérios previamente estabelecidos para a coleta de dados. Por exemplo, se você estiver entrevistando idosos e, nos critérios de inclusão, foi estabelecido não admitir pessoas com menos de 60 anos, descarte dados de pessoas com 59 anos ou menos, mesmo que os dados dessas pessoas sejam compatíveis com os demais.

Dado perdido: é o dado que existe, mas seu valor não está registrado.

Quando as amostras são grandes, alguns dados podem ser perdidos por *acaso*. Por exemplo, dados de alguns participantes podem se extraviar nos grandes levantamentos epidemiológicos sem que isso afete o resultado da pesquisa.

Não se pode, porém, considerar casual a perda de dados por razões associadas à natureza da pesquisa. No caso de questionários, se muitas pessoas não respondem a determinada pergunta, é preciso avaliar e discutir o que aconteceu. A pergunta estava redigida de forma clara? As pessoas

não queriam – por algum motivo – responder à pergunta? Qual seria esse motivo? Encontrar o que motivou a perda de dados ou, pelo menos, lançar uma hipótese sobre isso é muito importante.

Em ensaios clínicos, dados podem ser perdidos devido às próprias exigências da pesquisa. Por exemplo, se os participantes de pesquisa devem voltar com frequência ao centro de pesquisa para avaliações, dados de vários deles podem começar a faltar a partir de determinado momento – simplesmente porque eles se cansaram de cooperar.

Por outro lado, é preciso cuidado com as informações registradas apenas como “Sim” e “Não” quando se usam dados de arquivos, prontuários e fichas clínicas. Existe certa tendência a achar que dado não registrado significa “Não” em vez de considerar que o dado foi perdido. Então, se você estiver procurando informação sobre a presença de determinado sintoma, não entenda que ausência de registro significa ausência do sintoma. Procure saber o que realmente aconteceu.

Ainda, muitas vezes as datas são registradas apenas parcialmente. Por exemplo, escreve-se como data da primeira consulta: 08/1999. É razoável considerar que essa consulta foi no dia 15, só porque esse dia fica no meio do mês, em lugar de considerar que o dado foi perdido. Mas é preciso cuidado com outras datas que possam constar da ficha clínica e que tornem a data arbitrada impossível. Se a data de morte foi, por exemplo, 13/08/1999, ninguém conseguirá explicar a consulta.

Finalmente, é preciso cuidado quando se coletam dados para, a partir deles, calcular outra variável. Se você precisa calcular o índice de massa corporal (IMC) de seus pacientes, é melhor verificar se todos os dados de peso e altura foram registrados antes de começar os cálculos. Retire da amostra todos os participantes para os quais faltam informações ou encontre a informação perdida.

Dado censurado: é um dado que existe, mas não sabemos seu valor, embora saibamos que esse valor está além ou aquém de certo limite.[2] É comum aparecerem dados censurados nas seguintes situações:

1. Em ensaios clínicos de longo prazo, em que o resultado de interesse é o tempo de sobrevivência. É claro que o ensaio deve terminar em certo momento. O tempo de sobrevivência dos pacientes que continuaram vivos ao término do ensaio é um dado censurado.
2. Nas medidas obtidas por meio de aparelhos que têm limite mínimo para as medições. Nesses casos, o valor real de algumas unidades pode, eventualmente, estar abaixo do limite mínimo do aparelho. Esses dados são ditos censurados porque se sabe, apenas, que seus valores reais estão abaixo desse limite.
3. Nos ensaios em que a variável em análise é o tempo decorrido entre o início das observações e determinado desfecho. Se esse desfecho não ocorre no período de tempo estabelecido para a execução do ensaio para algumas unidades, os dados dessas unidades

são chamados de censurados.

Exemplo 1.10

Imagine que um pesquisador produziu inflamação em uma das patas de ratos de laboratório para depois verificar o tempo de recuperação deles quando tratados com diferentes anti-inflamatórios. Se o pesquisador medir as patas dos ratos no instante em que produziu a inflamação, 24 e 48 horas depois, alguns animais poderão permanecer com as patas inflamadas depois das 48 horas determinadas pelo protocolo para coleta de dados. Esses dados são ditos censurados porque, no período de tempo estabelecido para observação, o desfecho esperado não ocorreu.

1.4. DADOS UNIVARIADOS E DADOS BIVARIADOS

Os dados estatísticos são, muitas vezes, designados conforme o número de variáveis envolvidas na pesquisa.

Dados univariados: se, na pesquisa, for coletada uma única variável, dizemos que os dados são univariados. Por exemplo, se um dentista levantar o número de cáries em crianças da pré-escola, estará trabalhando com uma única variável. Seus dados serão, portanto, univariados.

Dados bivariados: quando, na pesquisa, são coletadas duas variáveis com a finalidade de estabelecer associações entre elas, dizemos que os dados são bivariados. Por exemplo, se um fisioterapeuta levantar dados de peso e altura de adolescentes do último ano do curso fundamental para calcular o índice de massa corporal, estará trabalhando com duas variáveis. Seus dados serão, portanto, bivariados.

RESUMO E OBJETIVO DO CAPÍTULO

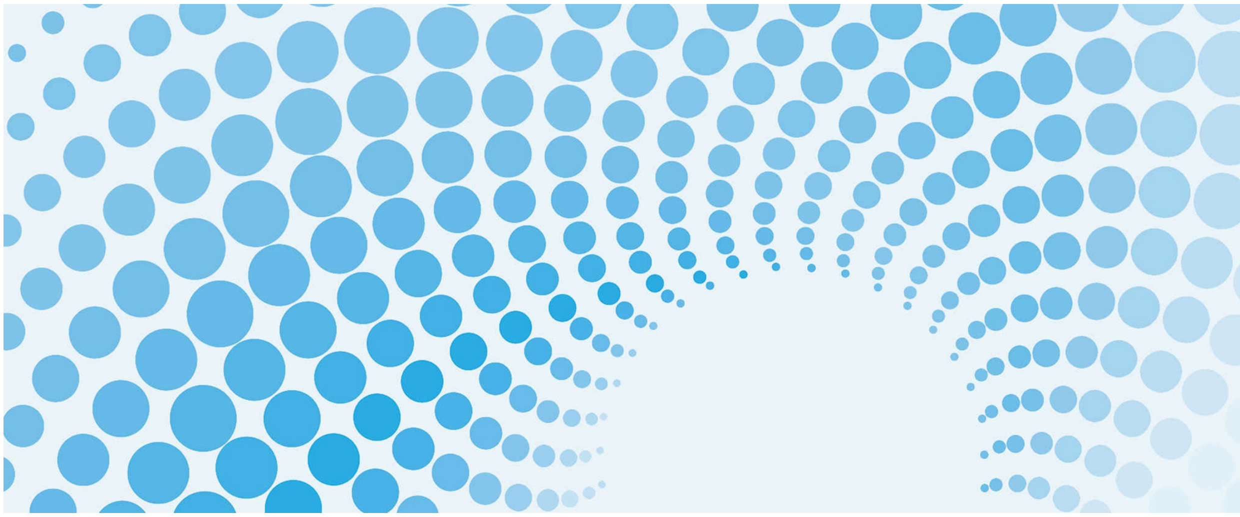
Lido este Capítulo, você deverá saber definir e exemplificar:

1. dado e variável
2. variável nominal
3. variável ordinal
4. variável discreta
5. variável numérica
6. dados perdidos
7. dados discrepantes
8. dados censurados
9. dados univariados e bivariados

---

[1] VIEIRA, S. Erros sistemáticos e aleatórios nas medições. Disponível em: https://www.blogger.com/blogger.

[2] Foi definida aqui a censura tipo I, mas existem outros tipos de censura. Veja, por exemplo: COX, D.R.; OAKES, D. *Analysis of Survival Data*. Londres: Chapmann & Hall, 1984.



1.5.1. Classifique as variáveis como nominais, ordinais, discretas ou contínuas: a) densidade óssea; b) marcas comerciais de um mesmo analgésico; c) comprimento de implantes; d) gosto (doce ou amargo); e) sexo; f) número de dentes permanentes irrompidos em uma criança; g) número de pacientes atendidos por dia em um consultório; h) altura da face; i) qualidade percebida do atendimento (bom, normal, ruim).

1.5.2. a) Dê um exemplo de situação em que o pesquisador pode atribuir nota ao que observa. b) Dê um exemplo em que o pesquisador obtém dados univariados.

1.5.3. Para mostrar que o uso de vocabulário especializado dificulta o entendimento do Termo de Consentimento Livre e Esclarecido (TCLE),[3] um pesquisador organizou um questionário para ser respondido pelos atendentes de um hospital. Havia questões sobre idade, sexo, escolaridade, doenças presentes, pressão arterial, medicamentos usados, alergia à aspirina e terminava perguntando se a pessoa estaria disposta a participar de um experimento para testar "a ação analgésica do ácido acetilsalicílico para eventuais cefaleias". Identifique os tipos de dados.

1.5.4. São dadas as notas de 16 alunos em uma prova. Você vê algum valor discrepante?

10; 9; 7; 8; 7; 9; 10; 7; 8; 8; 8; 9; 10; 2; 7; 9

1.5.5. Você pode solicitar às mães de crianças de 4 e 5 anos de idade que avaliem o medo que seus filhos têm de ir ao dentista usando uma escala visual analógica? Como isso pode ser planejado? Seria mais fácil usar outra forma de medir?

1.5.6. Reveja o exercício 1.5.3. Você acha que a última questão foi escrita de maneira adequada para que as atendentes de enfermagem possam responder? Ou, ao contrário, foi escrita para obter o número de não respondentes e/ou respostas inadequadas?

1.5.7. Para estudar o tempo de latência de um sonífero usando ratos de laboratório, um pesquisador administrou o sonífero a 10 ratos e registrou o tempo que cada um demorou para dormir: 2 demoraram meio minuto para dormir, 4 demoraram 1 minuto, 3 demoraram 1,5 minuto e 1 não dormiu durante o tempo estabelecido para obter os resultados do estudo. Algum dado foi censurado?

1.5.8. Um professor de educação física levantou dados dos alunos de uma escola pública: sexo, idade, peso, estatura, doenças, tratamento médico (em tratamento ou não) e esporte que pratica. Identifique o tipo de cada variável estudada.

1.5.9. Um homem é atendido na emergência de um hospital com queixa de dor no peito. O médico que o atende diagnostica angina e a classifica como severa. Essa classificação "severa" caracteriza uma variável:

a) numérica

b) nominal

c) ordinal

d) quantitativa

1.5.10. Responda: falso ou verdadeiro? "Um dos tratamentos de um ensaio clínico conduzido para comparar cinco tratamentos causa mais desconforto do que os outros. Provavelmente, esse tratamento terá mais dados censurados".

---

[3] Veja a Resolução CNS nº 196/1996 do Conselho Nacional de Saúde do Ministério da Saúde.

2

Testes Estatísticos

A leitura sistemática das revistas da área de saúde deixa evidente que os pesquisadores, embora trabalhem com *amostras, generalizam seus achados* para toda a *população* de onde a amostra foi retirada.[4] E, para generalizar seus achados, os pesquisadores aplicam *testes estatísticos*.

Vamos apresentar aqui a lógica desses testes por meio de um exemplo histórico. Em meados do século XIX, o cirurgião Joseph Lister (1827-1912), que estudava os trabalhos de Pasteur, percebeu que o estudo das bactérias e a prática da cirurgia eram ciências interdependentes.[5] Sendo isso verdade – Lister ponderou –, a assepsia das salas cirúrgicas deveria aumentar as taxas de sobrevivência nos pós-operatórios.

Para verificar essa hipótese, Lister fez um ensaio. Distribuiu 75 pacientes que iriam ser submetidos a uma cirurgia em dois grupos: o grupo controle, que foi submetido à cirurgia em salas nas condições usuais do hospital na época,[6] e o grupo com intervenção, que foi submetido à cirurgia em salas onde a assepsia havia sido feita com ácido fênico.

Lister tinha sua hipótese: a assepsia das salas cirúrgicas deveria aumentar a taxa de sobrevivência nos pós-operatórios. Outros cirurgiões tinham, porém, outra hipótese: a assepsia das salas cirúrgicas não teria qualquer efeito sobre as taxas de sobrevivência nos pós-operatórios. Os resultados do ensaio de Lister estão apresentados na Figura 2.1.

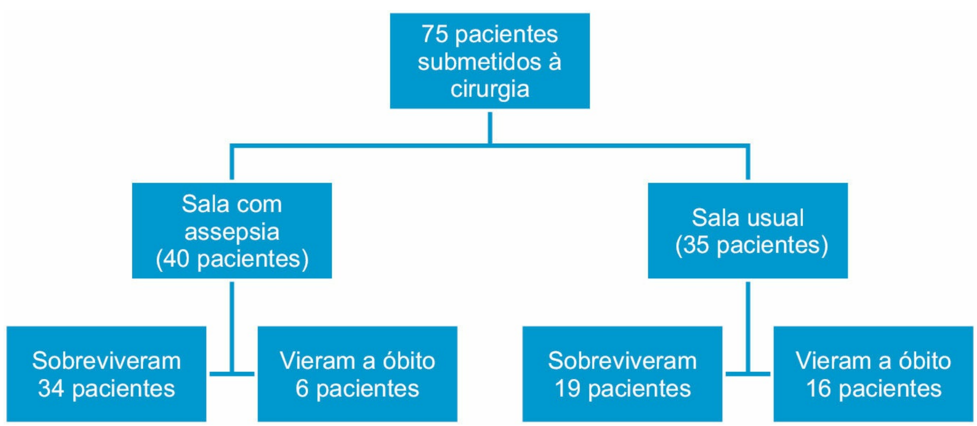


Figura 2.1 – Resultados do experimento de Lister.

A diferença das taxas de sobrevivência entre os dois grupos é grande. Veja: sobreviveram 34 dos 40 pacientes do grupo com intervenção. A taxa de sobrevivência é

$$\frac{34}{40} \times 100 = 85{,}0\%$$

Sobreviveram 19 dos 35 pacientes do grupo controle. A taxa de sobrevivência é

$$\frac{19}{35} \times 100 = 54{,}3\%$$

A diferença entre grupos é

$$85{,}0 - 54{,}3 = 30{,}7\%$$

Será que uma diferença entre grupos tão grande como essa foi suficiente para convencer os médicos da época da necessidade de assepsia? Não, como mostra a história da Medicina, mas como essa questão seria tratada hoje? Aplicando um teste estatístico.

2.1. DECIDINDO SOBRE AS HIPÓTESES

Vamos mostrar a construção de um teste estatístico usando o ensaio de Lister. Parece razoável considerar que há duas explicações possíveis para a diferença encontrada entre as taxas de sobrevivência:

1. *Acaso*. A diferença observada entre grupos seria obra do acaso.
2. *Assepsia*. A diferença observada entre grupos seria devida à intervenção (assepsia da sala cirúrgica).

Para decidir por uma das explicações, toda revista científica julga, hoje, mandatório aplicar um teste estatístico. Veja como se constrói o teste.

1. Comece considerando que a intervenção *não* tem efeito. Esta é a *hipótese da nulidade*[7] ("nulidade" porque diz que o efeito da intervenção é nulo). Indica-se por H_0 (lê-se agá zero). No exemplo, a hipótese é a de que a assepsia *não* tem efeito, ou seja, as taxas de sobrevivência *são iguais nos dois grupos*. Escreve-se:

   H_0: a intervenção *não tem* efeito.

2. Em seguida, considere a nova proposta: a intervenção *tem* efeito. Essa é a *hipótese alternativa.* Indica-se por H_1 (lê-se agá um). No exemplo, o pesquisador considerava que as taxas de sobrevivência *são maiores no grupo que recebeu a intervenção.* Escreve-se:

   H_1: a intervenção *tem* efeito positivo.

3. Agora, pergunte-se: se a *hipótese da nulidade for verdadeira,* é provável obter uma diferença entre taxas de sobrevivência tão grande quanto ou maior do que a que Lister obteve?
4. Calcule a probabilidade de ocorrer diferença entre taxas de sobrevivência tão grande ou maior do que a que Lister obteve, *considerando verdadeira a hipótese da nulidade*. Essa probabilidade[8] denomina-se *p-valor*.
5. Olhe o *p*-valor e admita:
   a) *p-valor pequeno* significa que é *pouco provável* obter uma *diferença tão grande ou maior* do que a observada se a *hipótese da nulidade for verdadeira.* Então, rejeite a hipótese da nulidade.
   b) *p-valor grande* significa que é *provável* obter *diferença tão grande ou maior* do que a observada se a *hipótese da nulidade for verdadeira.* Nesse caso, não rejeite a hipótese da nulidade.

Quão pequeno deve ser o *p*-valor para que você decida rejeitar a hipótese da nulidade? Por

tradição (não há nenhuma razão teórica para isso), o *p*-valor é considerado pequeno quando:

a) É menor do que 0,05. Escreve-se $p < 0{,}05$. Diz-se, então, que o resultado é *significante* e a hipótese da nulidade é rejeitada. Veja a Figura 2.2. Ou, então,

b) É menor do que 0,01. Escreve-se $p < 0{,}01$. Diz-se que o resultado é *altamente significante*. A hipótese da nulidade é rejeitada.

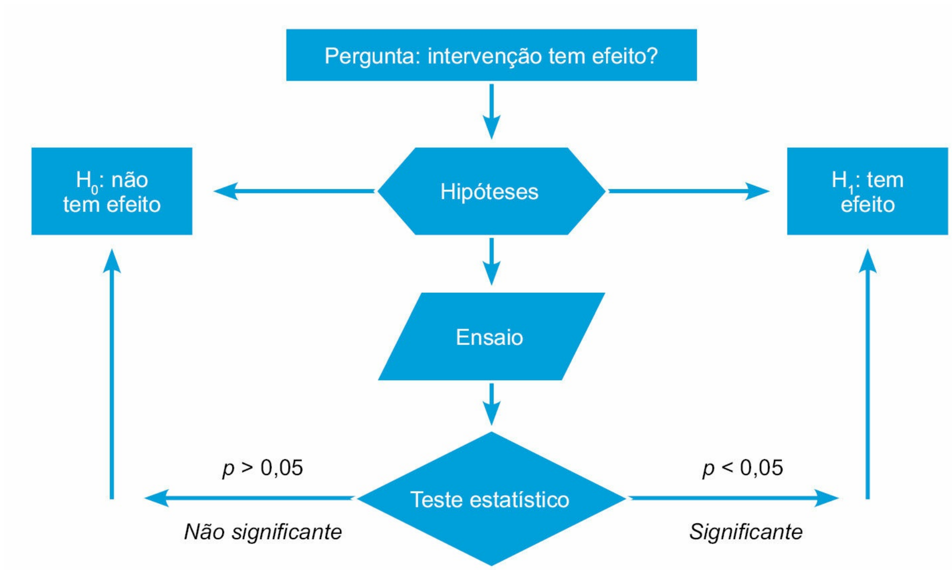


Figura 2.2 – Tomada de decisão: significante ou não significante?

Exemplo 2.1

Lister obteve os dados apresentados na Figura 2.1 . Sob a hipótese de que a assepsia não tem efeito sobre a taxa de sobrevivência dos operados (hipótese da nulidade), a probabilidade de ocorrer uma diferença igual ou maior do que a obtida por Lister é *p*-valor = 0,0024. Conclui-se então que a intervenção (assepsia) tem efeito.[9]

Sobrevivência de operados em salas cirúrgicas com e sem assepsia

Sobrevivência	**Assepsia na sala cirúrgica**		**Total**
	Sim	**Não**	
Sim	34	19	53
Não	6	16	22
Total	40	35	75
Taxa de sobrevivência	85,00%	54,30%	
p-valor	0,0024 < 0,05		

Fonte: WINSLOW, C. *The Conquest of Epidemic Disease*. Princeton: Princeton University Press, 1943. p. 303. Apud ALIAGA, M.; GUNDERSON, B. Interactive Statistics. 2nd ed. New Jersey: Prentice Hall, 2003. p. 673.

2.2. MEDINDO A INCERTEZA

2.2.1. Calculando o *p*-valor

Com base nos dados de uma amostra, os pesquisadores calculam o *p*-valor, que é a probabilidade de obter uma diferença tão grande ou maior do que a observada se o tratamento *não* tiver efeito.

Se o *p*-valor for pequeno, os pesquisadores fazem uma *inferência estatística:* concluem que o tratamento tem *efeito na população* de onde retiraram a amostra. No entanto, os pesquisadores não têm 100% de certeza de que essa conclusão está correta. Eles apenas sabem que a probabilidade de essa conclusão estar errada é muito pequena (é o *p*-valor).

Calcular o *p*-valor é difícil, mas hoje se usam computadores e programas prontos para fazer os cálculos. Antes da popularização dos computadores, usavam-se tabelas para determinar a significância dos resultados – e essas tabelas ainda são fornecidas nos livros de Estatística e na internet.

Não pense, porém, que ficou fácil aplicar um teste estatístico e obter o *p*-valor. É preciso, primeiramente, *escolher o teste* que será utilizado em função das hipóteses que se quer testar. Depois, é preciso verificar se a variável em análise atende às *pressuposições exigidas* para a aplicação do teste.[10]

E tenha muito cuidado na interpretação do *p*-valor,[11] que não é intuitiva porque usa a contradição: a hipótese colocada em teste afirma o contrário do que o pesquisador acha que é (ou gostaria que fosse) verdadeiro. Talvez por isso o *p*-valor seja, muitas vezes, mal interpretado. Para evitar que isso aconteça com você,[12] tenha sempre em mente que:

p-valor é a *probabilidade de obter uma diferença entre grupos tão grande ou maior do que a obtida* quando essa diferença não existe.

Em outras palavras:

p-valor é a probabilidade de o pesquisador estar errado quando diz que os grupos em comparação são diferentes.

Veja bem: o *p*-valor nada diz sobre o fato de H_0 *ser ou não ser verdadeiro*. Informa, apenas, que a decisão de rejeitar H_0 pode estar errada. No entanto, é o pesquisador quem deve resolver – juntando todas as informações que tem ao *p*-valor que obteve – se a hipótese da nulidade parece improvável o bastante para que ele corra o risco de dizer que *deve ser rejeitada*.

2.2.2. Apresentando o nível de significância

Para aplicar um teste estatístico, o pesquisador começa levantando a hipótese de que o tratamento *não* tem efeito. É a *hipótese da nulidade*, que se indica por H_0 (lê-se agá-zero). No entanto, o pesquisador acredita no contrário: acha que o novo tratamento tem efeito. Esta é a *hipótese alternativa* – que se indica por H_1 (lê-se agá-um).

O pesquisador faz o teste estatístico, que fornece o *p*-valor. Se o *p*-valor for pequeno – tradicionalmente, menor do que 0,05 –, o pesquisador *conclui que o tratamento tem efeito* na população de onde a amostra foi retirada. O pesquisador sabe que o *p*-valor é a probabilidade de essa conclusão estar errada.

Em teoria, o pesquisador deve estabelecer a probabilidade de errar quando diz que existe diferença entre grupos – e essa diferença não existe – antes de proceder ao teste. Essa probabilidade é o *nível de significância* do teste, que se indica por α (lê-se alfa). Quando você lê $p \leq 0{,}05$, saiba que a hipótese da nulidade (de que o tratamento não tem efeito) foi rejeitada no *nível de significância* de *5%*.

Teoricamente, a probabilidade de rejeitar H_0 quando H_0 é verdadeiro (o *nível de significância* α) deve ser estabelecida antes de o teste ser aplicado.

Exemplo 2.2

Lister considerou que a assepsia das salas cirúrgicas aumenta as taxas de sobrevivência nos pós-operatórios. Conduziu, então, um ensaio para mostrar isso. Hoje em dia, na publicação dessa pesquisa, obrigatoriamente estariam descritos:

- H_0: Na população, as taxas de sobrevivência são iguais nos dois grupos: tratado (com assepsia) e controle (sem assepsia).
- H_1: Na população, a taxa de sobrevivência no grupo tratado (com assepsia) é maior do que a taxa de sobrevivência no grupo controle (sem assepsia).
- Nível de significância: $\alpha = 0{,}05$.

Obtidos os dados, teria sido feito um teste estatístico para comparar as taxas de sobrevivência. Sob H_0, a probabilidade de ocorrer uma diferença igual ou maior do que a obtida por Lister é *p*-valor = 0,0024. Como o *p*-valor é menor do que o nível de significância adotado, isto é, $p < 0{,}05$, a conclusão é de que o ensaio trouxe evidência de que a assepsia nas salas de cirurgia aumenta a taxa de sobrevivência em cirurgias de amputação ($\alpha = 5\%$).

2.2.3. Avaliando o poder do teste

Poder[13] *do teste estatístico* é a probabilidade de o teste rejeitar a hipótese da nulidade quando a hipótese da nulidade é falsa.

Em outras palavras:

Poder do teste estatístico é a probabilidade de o teste detectar uma diferença que realmente existe na população.

Exemplo 2.3

Imagine que você está conduzindo uma série de ensaios com uma droga eficaz. Estão em comparação dois grupos: o tratado (que recebe a droga) e o controle (que não recebe a droga). Se você aplicar um teste estatístico com poder de 0,90 para comparar

grupos, deve esperar resultados estatisticamente significantes em 90% das vezes. Em 10% dos casos, não deve esperar resultados estatisticamente significantes. Então, poder do teste é a probabilidade de encontrar a diferença entre os dois grupos quando essa probabilidade realmente existe.

Vários fatores afetam o poder de um teste estatístico. Vamos mencionar alguns, imaginando que você esteja aplicando sempre o *mesmo teste.*[14]

- O poder do teste depende do *tamanho da diferença*: os testes estatísticos têm mais poder para detectar uma diferença de 40% do que uma diferença de 2%.
- O poder do teste depende do *nível de significância adotado.* No *nível de significância* de *5%*, a hipótese da nulidade será rejeitada se $p < 0{,}05$. *Diminuir o nível de significância* faz diminuir a probabilidade de rejeitar a hipótese da nulidade – seja ou não verdadeira – e, consequentemente, *diminui o poder do teste.*
- O poder do teste depende do *tamanho da amostra.* A confiança na informação aumenta quando aumenta a quantidade de dados.

IMPORTANTE: Quando se calcula o tamanho da amostra, é comum adotar – embora não haja qualquer justificativa teórica para isso – nível de significância de 5% e poder de teste de 80%. Isso significa que você admite até 5% de probabilidade de estar errado ao dizer que os grupos são diferentes e 80% de probabilidade de detectar uma diferença que realmente existe.

2.3. TESTES UNILATERAIS E BILATERAIS

Para aplicar um teste de hipóteses é preciso, primeiramente, decidir se o teste será unilateral ou bilateral.

Um teste é *unilateral* quando a hipótese da *nulidade* só pode ser rejeitada se a diferença entre grupos tiver o sentido especificado pelo pesquisador (p. ex., o pesquisador diz, no projeto, que espera efeito positivo do novo tratamento).

Exemplo 2.4

Foi apresentado o resultado do teste estatístico para os dados de Lister. Veja as hipóteses:

- H_0: assepsia *não tem efeito* sobre a taxa de sobrevivência dos operados.
- H_1: assepsia *aumenta* a taxa de sobrevivência dos operados.

O teste unilateral se justifica: é razoável verificar se os dados apresentados por Lister já mostravam o que se sabe hoje.

Um teste é *bilateral* quando a hipótese da *nulidade* é rejeitada *qualquer que seja o sentido (o sinal) da diferença entre grupos*.

Exemplo 2.5

Dez ratos machos adultos, criados em laboratório, foram separados aleatoriamente em dois grupos: o grupo controle, que recebeu a ração normalmente usada no laboratório, e o grupo tratado, que recebeu uma ração experimental. Como não se sabe o sentido da diferença, isto é, se a nova ração vai aumentar ou diminuir o peso dos ratos, parece razoável fazer um teste bilateral:

- H_0: em média, o peso dos ratos tratados com a nova ração *é igual* ao peso dos ratos tratados com a ração conhecida.
- H_1: a média dos pesos dos ratos que receberam a nova ração é estatisticamente *diferente* da média dos ratos tratados com a ração conhecida.

Os estatísticos em geral discordam da ideia de testes unilaterais, que seriam válidos somente quando fosse *impossível* ocorrer uma diferença em sentido contrário ao esperado. Veja as razões por que eles discordam da aplicação de testes unilaterais:

1. Os testes bilaterais são *mais seguros* – porque o tratamento pode dar resultado contrário ao esperado.
2. Os testes bilaterais são *mais conservadores*, isto é, têm menor probabilidade de rejeitar H_0. Como grande parte dos experimentos não satisfaz às exigências metodológicas nem às pressuposições exigidas para aplicar os testes estatísticos, é melhor trabalhar com testes que têm menor probabilidade de detectar significância.

Alguns pesquisadores preferem o teste unilateral, que é referido como *teste de superioridade*. O argumento em favor dos testes unilaterais é o de que uma boa revisão da literatura mostra o sentido da diferença. No entanto, a situação será constrangedora se a diferença for significante, mas em sentido contrário ao esperado.

2.4. TESTES PARAMÉTRICOS E NÃO PARAMÉTRICOS

Na área médica, são muito usados o teste *t* de Student,[15] a análise de variância e o teste de Tukey.[16] Tais testes exigem, para sua aplicação, que a variável em análise seja *numérica* e as hipóteses sejam feitas sobre *parâmetros*, como médias e variâncias.[17] Daí o nome: *testes paramétricos*. Por exemplo, para comparar a altura de meninos e meninas da mesma idade, o pesquisador pode optar por testes paramétricos – e testar as diferenças de médias (pelo teste *t* de Student) e de variâncias (pelo teste *F* de Snedecor).

Os testes paramétricos têm, ainda, outras exigências. Por exemplo, o teste *t* exige pressupor que a variável em análise tenha distribuição normal ou, pelo menos, simétrica. O teste *F* exige, além da pressuposição de variável com distribuição normal ou aproximadamente normal, ou pelo menos simétrica, a pressuposição de homocedasticia (homogeneidade de variâncias, ou seja, a mesma variabilidade dentro de grupos).

E o que faz o pesquisador quando verifica que a variável que estuda não atende às pressuposições exigidas para procedimento de um teste paramétrico? Aplica um teste não paramétrico. Esses testes não exigem que a variável em análise seja numérica e não exigem pressuposição a respeito da distribuição da variável. *Não pense*, porém, que pode "substituir" um teste paramétrico por um não paramétrico: as hipóteses são diferentes. Quando você aplica um teste *t* de Student (paramétrico), a hipótese da nulidade é de que *as médias populacionais são iguais*. E quando você aplica um teste de Mann-Whitney (não paramétrico), a hipótese da nulidade é de *que as distribuições são iguais.*[18]

Exemplo 2.6

Imagine que um pesquisador quer verificar se a dipirona é mais eficaz do que o paracetamol no controle da dor após determinada intervenção odontológica. A intensidade da dor pode ser registrada por meio de notas, atribuindo valor 0 para nenhuma dor e valor 5 para dor intensa. Como a intensidade de dor é variável *ordinal*, *não tem sentido* calcular médias e, portanto, *não tem sentido* aplicar um teste paramétrico. O pesquisador deve buscar um teste não paramétrico.

A lógica dos testes não paramétricos é simples e aplicá-los é bem mais fácil do que aplicar seus equivalentes paramétricos. No entanto, como tudo na vida tem um preço, você não vai se surpreender com o fato de os testes não paramétricos terem algumas desvantagens:

- Os testes paramétricos são mais *poderosos*, ou seja, têm maior probabilidade de rejeitar a hipótese da nulidade quando essa hipótese é falsa.
- Os testes paramétricos são bastante *robustos*, ou seja, pequenas transgressões às pressuposições exigidas não invalidam os resultados – principalmente se as amostras forem de tamanho grande ou moderado.

Portanto, para a análise de dados de variáveis com distribuição aproximadamente normal obtidas de grandes amostras, a melhor opção ainda é um teste paramétrico, mais poderoso –

mesmo que algumas pressuposições não estejam totalmente satisfeitas. Se você optar por um teste não paramétrico, estará optando por um teste com *menos poder*.

De qualquer forma, diante de um problema real de pesquisa, você precisa se decidir por um teste paramétrico ou por um teste não paramétrico. Aplique um teste paramétrico se as seguintes afirmativas forem verdadeiras:

1. Dados quantitativos.
2. As pressuposições exigidas para a aplicação do teste escolhido estão satisfeitas ou são pequenas as transgressões.
3. Amostra é de tamanho grande ou, pelo menos, moderado.

Aplique um teste não paramétrico quando estiver diante de uma das seguintes situações:

1. Dados qualitativos.
2. As pressuposições exigidas para a aplicação de testes paramétricos não estão satisfeitas.
3. A amostra é pequena.
4. Existem dados discrepantes ou censurados, que podem tornar mais indicado calcular medianas – e não as médias.

2.4.1. Algumas indicações

Testes para comparar dois grupos independentes:

- Para a análise de variável numérica com distribuição normal ou, pelo menos, simétrica e amostra grande (maior do que 30), aplique um teste paramétrico: o teste t de Student.[19]
- Para a análise de variável ordinal ou de variável numérica com distribuição não simétrica e/ou amostra muito pequena, aplique teste não paramétrico: o teste de Mann-Whitney ou o teste da mediana, que serão tratados no Capítulo 3 deste livro.
- Para a análise de variável nominal, use o teste de χ^2 de Pearson ou o teste exato de Fisher, que serão tratados no Capítulo 5 deste livro.

Testes para comparar dois grupos dependentes:

- Para a análise de variável numérica com distribuição normal ou, pelo menos, simétrica e/ou uma amostra bastante grande, aplique um teste paramétrico: o teste t de Student para amostras pareadas.[20]
- Para a análise de variável ordinal ou de variável numérica com distribuição não simétrica e/ou amostra muito pequena, aplique um teste não paramétrico: o teste de Wilcoxon ou o teste do sinal, que serão tratados no Capítulo 4 deste livro.
- Para a análise de variável nominal, use o teste de χ^2 de Mc-Nemar,[21] tratado no Capítulo 5 deste livro.

Testes para comparar mais de dois grupos independentes:

- Se a variável for numérica e tiver distribuição normal ou aproximadamente normal, ou tiver

pelo menos distribuição simétrica e a amostra for bastante grande, faça uma análise de variância[22].

- Se a variável for ordinal ou numérica, mas a distribuição não for simétrica e/ou a amostra for pequena, aplique um teste não paramétrico: o teste de Kruskal-Wallis ou o teste da mediana, mostrados no Capítulo 4 deste livro.

Testes para comparar mais de dois grupos dependentes:

- Se a variável for numérica e tiver distribuição normal ou aproximadamente normal, ou tiver pelo menos distribuição simétrica e a amostra for bastante grande, faça uma análise de variância.[23]
- Se a variável for ordinal ou numérica, mas a distribuição não for simétrica e/ou a amostra for pequena, aplique um teste não paramétrico: o teste de Friedman, que veremos no Capítulo 4 deste livro.

RESUMO E OBJETIVO DO CAPÍTULO

Neste capítulo foi apresentada a lógica de um teste estatístico. São definidos *p*-valor, testes unilaterais e bilaterais, testes paramétricos e não paramétricos. Portanto, uma vez lido este capítulo, você deverá saber o que significam:

1. Hipóteses estatísticas.
2. *p*-valor e nível de significância.
3. Poder do teste.
4. Teste unilateral.
5. Teste bilateral.
6. Testes paramétricos e não paramétricos.

---

[4] Para estudar amostragem, veja: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016.

[5] DE PAOLLO, C.; PASTEUR, Lister: a chronicle of scientific dependence. The Victorian Web: Literature, History and Culture in the age of Victoria. Disponível em: http://www.victorianweb.org/science/health/depaolo.html. Acesso em 23 de outubro de 2017.

[6] Meados do século XIX.

[7] A maioria dos estatísticos brasileiros diz “hipótese nula”, traduzindo assim, de maneira a meu ver equivocada, a expressão inglesa *null hypothesis*.

[8] É difícil calcular essa probabilidade, mas esse cálculo é feito em computador.

[9] Foi feito um teste unilateral para comparar proporções.

[10] Nos próximos capítulos, veremos a indicação de vários testes e as fórmulas de cálculo. Usaremos tabelas disponíveis no final deste livro e mostraremos exemplos de resultados obtidos por meio de programas para computador.

[11] Esta argumentação está totalmente baseada em Motulsky, H. *Intuitive Statistics*. New York: Oxford University Press, 1995.

[12] Glantz, S. A. *Primer of Biostatistics*. 2nd ed. New York: McGraw, 1987. p. 97.

[13] Alguns autores dizem potência do teste (tradução de *power of a statistical test*).

[14] Alguns testes têm mais poder do que outros quando aplicados ao mesmo conjunto de dados.

[15] Veja, por exemplo: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Campus-Elsevier, 2016.

[16] Veja, por exemplo: VIEIRA, S. *Análise de variância*. São Paulo: Atlas. *E-book*.

[17] Veja, por exemplo: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016.

18 Veja a seção 5.2 do Capítulo 5 deste livro.

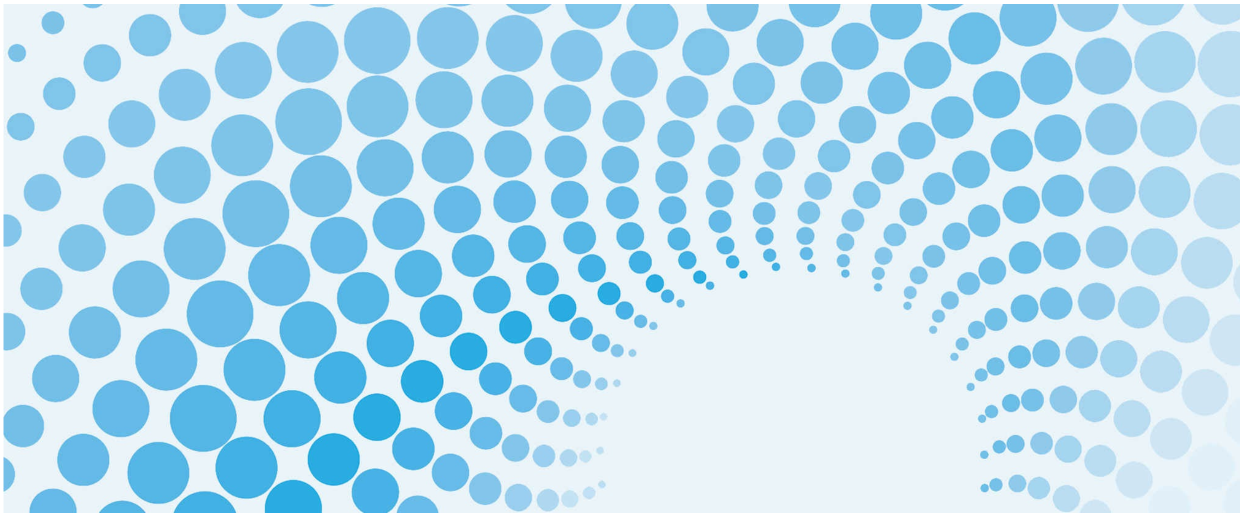
[19] Veja o procedimento para o teste *t* de Student em VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016. Capítulo 13.

[20] Id., ibid.

[21] Veja o Capítulo 5 deste livro.

[22] Veja o procedimento para análise de variância em VIEIRA, S. *Análise de variância*. São Paulo: Atlas, 2006.

[23] Id., ibid.



2.5.1. Construa as hipóteses no caso de pesquisa com o objetivo de: a) verificar se uma nova droga é melhor do que a tradicional; b) verificar se determinada dieta aumenta a longevidade; c) verificar se um produto é cancerígeno; d) verificar se uma vitamina aumenta o desempenho de atletas.

2.5.2. Para comparar a prevalência de alcoolismo entre homens e mulheres residentes do Rio de Janeiro, foram levantados os dados apresentados na Tabela 2.2. Construa as hipóteses. Foi feito o teste estatístico denominado qui-quadrado. O *p*-valor é 0,0004. O que você conclui?

Prevalência de alcoolismo segundo o sexo

Sexo	**Alcoolismo**		**Total**
	Sim	**Não**	
Masculino	29	559	588
Feminino	15	856	871
Total	44	1.415	1.459

Fonte: ALMEIDA, LM, COUTINHO, ESF. Prevalência de consumo de bebidas alcoólicas e de alcoolismo em uma região metropolitana do Brasil. Rev Saúde Pública 27 (1), 1993.

2.5.3. Um pesquisador pretende conduzir um experimento para testar uma hipótese[24]. Se resolver duplicar o tamanho da amostra, qual dos seguintes itens irá aumentar?

a) O poder do teste.

b) O efeito do tratamento.

c) A probabilidade de erro Tipo II.

Escolha a resposta:

1. O poder do teste.
2. O efeito do tratamento.
3. A probabilidade de erro Tipo II.
4. Todos os itens listados.
5. Nenhuma das alternativas listadas.

2.5.4. Para verificar se meninos e meninas da mesma idade têm igual velocidade de leitura e igual capacidade de interpretação de textos, um pesquisador pediu a um grupo de crianças que lessem determinada poesia. Para cada criança, cronometrou o tempo de leitura e atribuiu um conceito para medir a capacidade de interpretação. Você aplicaria um teste paramétrico para analisar qual das variáveis?

2.5.5. Você quer testar a hipótese de que uma moeda é bem balanceada. Quais são as hipóteses em teste? Você resolve jogar a moeda seis vezes e dizer que não é bem balanceada se aparecerem seis caras. Um estatístico calcula que a probabilidade de isso acontecer, quando a moeda é bem balanceada,[25] é

$$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=0{,}015625$$

Qual é o *p*-valor?

2.5.6. Dos 100 pacientes submetidos a determinada cirurgia, oito vieram a óbito. Que população você definiria para essa

estatística?

2.5.7. Poder do teste significa:[26]

a) Não rejeitar H_0 quando H_0 é verdadeira.

b) Não rejeitar H_0 quando H_0 é falsa.

c) Não rejeitar H_0 independentemente de H_0 ser falsa ou verdadeira.

d) Rejeitar H_0 quando H_0 é verdadeira.

e) Rejeitar H_0 quando H_0 é falsa.

2.5.8. Foi feito um ensaio para comparar o efeito analgésico de duas drogas, A e B, e aplicado um teste estatístico para comparar os resultados. Quais são as hipóteses em teste?

2.5.9. Foi medido o nível de determinado hormônio em 50 mulheres jovens que não estavam em gestação e em 50 mulheres jovens que estavam no primeiro trimestre de gestação. Foram obtidas as médias, 93 e 110, respectivamente, e obtido um *p*-valor para um teste unilateral igual a 0,001, porque havia informações na literatura de que gestantes teriam maior quantidade desse hormônio. Escreva as hipóteses e interprete o *p*-valor.

2.5.10. Escolha uma das alternativas abaixo para interpretar a expressão “resultado significante”:

a) A pesquisa é de boa qualidade.

b) O pesquisador aceitou a hipótese da nulidade com probabilidade de 0,5.

c) A probabilidade de o estudo ser verdadeiro é maior do que 5%.

d) A probabilidade de o pesquisador ter obtido o resultado que obteve por acaso é ⌊0,05.

2.5.11. Quais dos seguintes procedimentos reduz o poder de um teste de hipóteses[27]?

a) Aumentar o tamanho da amostra.

b) Aumentar o nível de significância.

c) Aumentar a probabilidade de erro Tipo II.

Escolha a resposta:

a) Aumentar o tamanho da amostra.

b) Aumentar o nível de significância.

c) Aumentar a probabilidade de erro Tipo II.

d) Todos os itens listados.

e) Nenhum dos itens listados.

---

[24] *Power of a Hypothesis Test*. Disponível em: http://stattrek.com/hypothesis-test/power-of-test.aspx?Tutorial. Acesso em 24 de outubro de 2017.

[25] Veja o Capítulo 9 de VIEIRA, S. *Introdução à Bioestatística*. 4. ed. Rio de Janeiro: Campus/Elsevier, 2008.

[26] Prova: Analista do CNMP (Conselho Nacional do Ministério Público), 2015.

[27] *Power of a Hypothesis Test*. Disponível em: http://stattrek.com/hypothesis-test/power-of-test.aspx?Tutorial. Acesso em 24 de outubro de 2017.

3

Testes não Paramétricos para Comparar Dois Grupos

Neste Capítulo são apresentados testes não paramétricos indicados para comparar dois grupos independentes ou dependentes, nos casos em que a variável em análise é ordinal ou é numérica.

3.1. POSTOS EM LUGAR DE DADOS

Os estudos de estatística se iniciam com os cálculos de médias e desvios padrão de dados coletados. Alguns dos testes não paramétricos mais conhecidos são, porém, aplicados *não* aos dados coletados, mas aos *postos* dos dados. Embora muitas pessoas aprendam a fazer esses testes (ou, simplesmente, utilizem um programa de computador), ainda é difícil deixar de olhar para as médias dos dados originais. No entanto, alguns dos testes que vamos ver neste Capítulo, e no Capítulo 4, fazem inferência com base nos *postos* dos dados – e são os postos que devem ser comparados.

Vamos ver como se obtêm os postos. Comece colocando os n dados observados em ordem crescente. Depois, atribua um número de ordem a cada dado observado. Esse número é o *posto*,[28] indicado por R. O menor posto é 1 e o maior posto é n.

Exemplo 3.1

Logo após cirurgia para extração de um dente, oito pessoas deram notas de 0 a 10 à atenção que receberam do cirurgião-dentista. As notas foram: 7,5; 6; 5,5; 8; 9,5; 6,5; 4,5; 9. Para conferir postos, coloque as notas em ordem crescente. Depois atribua um número de ordem a cada nota. Esse número é o *posto*.

Posto, segundo a nota conferida em uma pesquisa de satisfação

Nota	Posto
4,5	1
5,5	2
6	3
6,5	4
7,5	5
8	6
9	7
9,5	8

Podem ocorrer dados com o mesmo valor. Chamamos isso de *empate*.[29] Valores empatados devem receber o mesmo posto, mas faça assim: escreva os dados em ordem crescente e atribua aos dados, mesmo que empatados, postos diferentes. Depois, calcule a *média dos postos ocupados pelos valores empatados*. Essa média será o posto dos valores empatados.

Exemplo 3.2

Logo após uma cirurgia para extração de dente, oito pessoas deram notas de 0 a 10 à atenção que receberam do cirurgião-dentista: 7,5; 6; 5,5; 8; 9,5; 6; 4,5; 9. Para conferir postos às notas, coloque as notas em ordem crescente e atribua um número de ordem a cada nota. As notas em negrito constituem empate. Elas receberam postos 3 e 4. A média desses números é 3,5. Este é o posto desses dois empates.

Posto, segundo a nota conferida em uma pesquisa de satisfação

Dado	Posto
4,5	1
5,5	2
6,0	**3,5**
6,0	**3,5**
7,5	5
8,0	6
9,0	7
9,5	8

Se o número de empates for pequeno, a solução apresentada aqui para tratar os empates é satisfatória.[30]

Dica: se você estiver fazendo cálculos à mão, ou se a amostra for grande e os empates forem muitos, verifique as contas. É fácil: haja ou não empates, você terá sempre:

$$\sum R = n \times \frac{n+1}{2} \tag{3.1}$$

Exemplo 3.3

Para o Exemplo 3.2, a soma dos postos é:

$$\sum R = 1 + 2 + 3{,}5 + 3{,}5 + 4 + 5 + 6 + 7 + 8 = 36$$

Como $n = 8$:

$$n \times \frac{n+1}{2} = 8 \times \frac{8+1}{2} = 36$$

Para os dados do Exemplo 3.2, a igualdade (3.1) está verificada.

3.2. COMPARAÇÃO DE DOIS GRUPOS INDEPENDENTES

3.2.1. Grupos independentes

Dois grupos de dados são independentes se as unidades de um deles *não* têm relação com as unidades do outro.

Exemplo 3.4

Para testar a eficácia de uma nova droga para reduzir a pressão arterial, uma indústria farmacêutica fez o seguinte ensaio: recrutou voluntários que tinham pressão arterial alta e mediu a pressão arterial de cada um deles no início do ensaio. Depois os dividiu aleatoriamente em dois grupos. Um grupo, denominado tratado, recebeu a droga em teste e o outro grupo, denominado controle, recebeu o tratamento convencional. No final do ensaio, a pressão arterial de cada voluntário foi novamente medida. Calculou-se, então, a diferença entre a pressão no final do ensaio e a pressão no início, o que resultou em dois grupos independentes de dados.

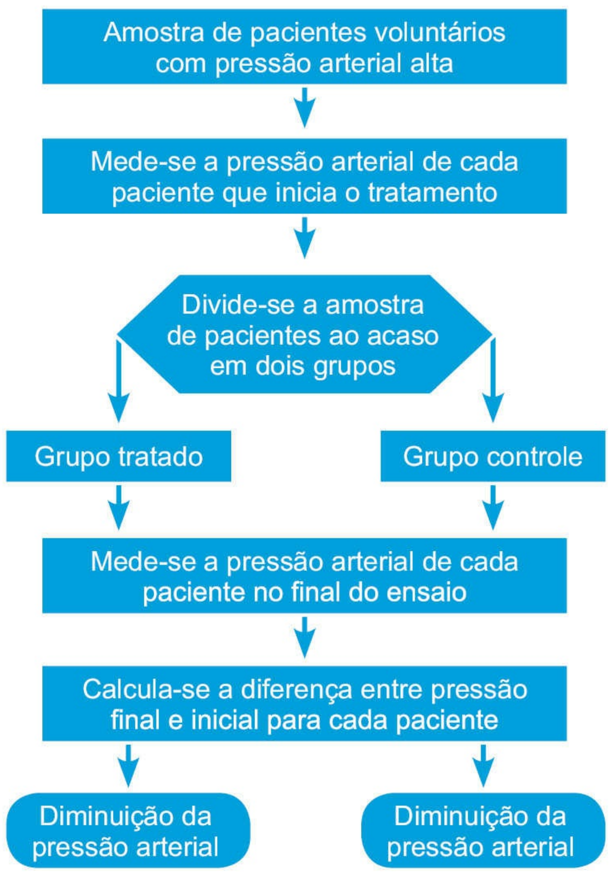


Figura 3.1 – Dados independentes.

3.2.2. Teste de Mann-Whitney

O teste de Mann-Whitney é também conhecido como teste *U* de Mann-Whitney, teste de Wilcoxon-Mann-Whitney, teste dos postos somados de Wilcoxon.

Indicação: O teste de Mann-Whitney é indicado para comparar dois grupos independentes, como homens e mulheres, tratado e controle, empregado e desempregado. Recomenda-se que as amostras tenham tamanhos $n_1 \geq 5$ e $n_2 \geq 5$. A variável em análise deve ser ordinal (como escala visual analógica ou itens de uma escala Likert) ou contínua (como peso, tempo, QI). As observações devem ser independentes, isto é, as unidades de um grupo não devem estar relacionadas com as unidades do outro grupo nem devem estar relacionadas entre si. A hipótese em teste é a de que postos dos dois grupos têm a mesma distribuição.[31] Portanto, o teste de Mann-Whitney *não* testa a hipótese de igualdade de médias da variável em estudo.

Para fazer o teste de Mann-Whitney e comparar dois grupos, 1 e 2, com n_1 e n_2 de dados, respectivamente:

Primeiro passo: Estabeleça as hipóteses e o nível de significância.

- H_0: os postos dos dois grupos têm a mesma distribuição.
- H_1: os postos dos dois grupos têm distribuições diferentes.

Segundo passo: Denomine Grupo 1 o grupo com o *menor número* de dados. Se $n_1 = n_2$, denomine Grupo 1 o grupo com a *menor soma* de postos. Faça $n_1 + n_2 = n$.

Terceiro passo: Junte os dois grupos em um só conjunto de dados, sem perder a identificação dos grupos. Ordene os dados e atribua postos a eles. Se houver empates, atribua aos valores iguais a média de seus postos, como já foi visto.

Quarto passo: Calcule a soma dos postos atribuídos aos dados do Grupo 1, isto é, õR_1; calcule a soma dos postos atribuídos aos dados do Grupo 2, isto é, õR_2.Verifique:

$$\sum R_1 + \sum R_2 = n \times \frac{n+1}{2} \tag{3.2}$$

Quinto passo: Calcule a *estatística U de Mann-Whitney*:

$$U = \sum R_1 - \left[\frac{n_1(n_1+1)}{2}\right] \tag{3.3}$$

Sexto passo: Se a amostra for de tamanho $n > 30$ e houver poucos empates, calcule:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \tag{3.4}$$

Sétimo passo: Compare o valor calculado de *z* com o valor crítico dado na tabela de distribuição normal padronizada (Tabela 1 do Apêndice), no nível estabelecido de significância.[32] Rejeite a hipótese da nulidade toda vez que o valor calculado de z for igual ou maior do que o valor crítico. Se estiver usando um programa de computador, você obterá o *p*-valor associado ao menor valor da estatística *U*.

Oitavo passo: Compare as medianas (ou as médias) dos *postos* dos dois grupos.

Exemplo 3.5

Foi conduzido um ensaio clínico para avaliar a eficácia de uma nova terapia antirretroviral para pacientes com HIV.[33] Trinta pacientes foram divididos aleatoriamente em dois grupos de mesmo tamanho: o controle, que recebeu terapia antirretroviral padrão, e o tratado, que recebeu a nova terapia antirretroviral. Os dois grupos foram monitorados por 3 meses. O resultado primário é a carga viral, relatada como número de cópias de HIV por mililitro de sangue. Os dados estão na Tabela 3.1.

Tabela 3.1 – Carga viral de pacientes com HIV segundo o grupo

Grupo	
Terapia padrão	**Nova terapia**
7.500	400
8.000	250
2.000	800
550	1.400
1.250	8.000
1.000	7.400
2.250	1.020
6.800	6.000
3.400	920
6.300	1.420
9.100	2.700
970	4.200
1.040	5.200
670	4.100
400	(1)

Dado censurado: foi obtido, mas o valor estava abaixo do limite mínimo para as medições do aparelho.

Primeiro passo

- H_0: os postos dos dois grupos têm a mesma distribuição.

• H_1: os postos dos dois grupos têm distribuições diferentes.

• Nível de significância α = 0,05.

Segundo passo: Denomine como Grupo 1 o grupo que recebeu a nova terapia e como Grupo 2 o que recebeu a terapia padrão, porque $n_1 = 14 < n_2 = 15$.

Terceiro passo: Atribua postos aos dados.

Tabela 3.2 – Carga viral de pacientes com HIV segundo o grupo (Tabela auxiliar)

Grupo			
1 (Nova terapia)		**2 (Terapia padrão)**	
Dado	**Posto**	**Dado**	**Posto**
250	1		
400	2,5	400	2,5
		550	4
		670	5
800	6		
920	7		
		970	8
		1.000	9
1.020	10		
		1.040	11
		1.250	12
1.400	13		
1.420	14		
		2.000	15
		2.250	16
2.700	17		
		3.400	18
4.100	19		
4.200	20		
5.200	21		
6.000	22		
		6.300	23
		6.800	24
7.400	25		
		7.500	26
8.000	27,5	8.000	27,5
		9.100	29
	205		230

Quarto passo: Calcule a soma dos postos atribuídos aos dados dos dois grupos e verifique os cálculos usando a fórmula (3.2):

$$\sum R_1 + \sum R_2 = n \times \frac{n+1}{2} =$$

$$205 + 230 = 29 \times \frac{29+1}{2} = 435$$

Quinto passo: Calcule a *estatística U de Mann-Whitney;* uma vez que $n_1 < n_2$ e $R_1 < R_2$. Use a fórmula (3.3):

$$U = \sum R_1 - \left[\frac{n_1(n_1+1)}{2}\right]$$

$$U = 205 - \left[\frac{14(14+1)}{2}\right] = 205 - 105 = 100$$

Sexto passo: A amostra tem tamanho 29, com poucos empates. Na verdade a amostra deveria ser maior do que 30. Estamos no limite. Mesmo assim, vamos usar a estatística *z*, com distribuição que se aproxima da normal padronizada. Então, calcule:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

$$z = \frac{100 - \frac{14 \times 15}{2}}{\sqrt{\frac{14 \times 15 \times (14 + 15 + 1)}{12}}} = -\frac{5}{\sqrt{525}} = -0{,}218218$$

Sétimo passo: Na tabela de distribuição normal padronizada, no nível de significância de 5% para um teste bilateral, encontra-se o valor crítico 1,959964. O *p*-valor é 0,827259. Portanto, não se rejeita a hipótese de que não há diferença entre as duas terapias.

Empates – Fique atento aos empates. Considere:

1. Se a amostra for grande, um número moderado de empates não muda muito o resultado, principalmente se os empates ocorrerem no mesmo grupo. Pode haver perda de poder estatístico se os empates ocorrerem em grupos distintos.
2. Quando os empates são muitos, é preciso fazer uma correção na fórmula[34] que dá o valor de *U*. A fórmula com correção não é dada neste livro, mas os programas para computador fazem, automaticamente, a correção.

Amostras pequenas: Quando as amostras são pequenas, a distribuição da variável *z* não se aproxima, satisfatoriamente, da distribuição normal padronizada. Se $n_1 \leq 15$ e $n_2 \leq 15$, não use, na sua pesquisa, a aproximação normal, ou seja, não calcule o valor de *z* como indicado na fórmula (3.4) do sétimo passo. Use a Tabela 3 do Apêndice, que dá os valores críticos de õR_1 para alguns níveis de significância. Veja como se usa essa tabela.

a) Se a hipótese alternativa (H_1) for a de que a distribuição dos dados no *grupo 1 é diferente da distribuição dos dados no Grupo 2*, encontre, na Tabela 3 do Apêndice, o par de valores

que está na coluna encabeçada com nível de significância $\alpha/2$. Por exemplo, se você adotou o nível de significância de 0,05, busque a coluna correspondente a 0,025. Rejeite a hipótese da nulidade se ΣR_1 for *igual ou menor* do que o *menor número* ou *igual ou maior* do que o *maior número do par*.

b) Se a hipótese alternativa (H_1) for a de que *a soma dos postos do grupo 1 é menor do que o Grupo 2*, encontre, na Tabela 3 do Apêndice, o par de valores que está na coluna correspondente ao nível de significância α. Rejeite a hipótese da nulidade se ΣR_1 for *igual ou menor* do que o *menor número do par*.

c) Se a hipótese alternativa (H_1) for a de que *a soma dos postos do Grupo 1 é maior do que o Grupo 2*, encontre, na Tabela 3 do Apêndice, o par de valores que está na coluna correspondente ao nível de significância α. Rejeite a hipótese da nulidade se ΣR_1 for *igual ou maior* do que o *maior número do par*.

Exemplo 3.6

Para definir a temperatura ambiente nos laboratórios de uma instituição, foram entrevistados nove técnicos de um dos laboratórios: quatro homens e cinco mulheres. Pediu-se a essas pessoas que apontassem a temperatura mais confortável, mas alguém levantou a hipótese de que homens apreciam temperaturas mais baixas. Os dados foram, então, anotados separadamente para cada sexo. Pressupondo que as respostas sejam de uma amostra representativa da população de técnicos dessa instituição, você diria que a temperatura mais confortável para homens tem a mesma distribuição da temperatura mais confortável para mulheres? Para responder a essa pergunta, é preciso fazer um teste estatístico – e o teste indicado aqui é o de Mann-Whitney – porque os dois grupos são independentes e a amostra é muito pequena.

Temperatura ambiente em graus centígrados, definida como mais confortável segundo o sexo

Sexo	
Homem	**Mulher**
20	24
18	26
23	21
19	22
	25

Primeiro passo

- Hipótese da nulidade: a variável tem igual distribuição nos dois grupos.
- Hipótese alternativa: homens preferem temperaturas mais baixas.
- Nível de significância $\alpha = 0,05$.

Segundo passo: Denomine como Grupo 1 o grupo de homens e Grupo 2 o grupo de mulheres, porque $n_1 = 4 < n_2 = 5$.

Terceiro passo: Atribua postos aos dados. Não há empates.

Temperatura ambiente em graus centígrados, definida como mais confortável segundo o sexo (Tabela auxiliar: postos)

Sexo

Homem		Mulher	
Temperatura	Posto	Temperatura	Posto
20	3	24	7
18	1	26	9
23	6	21	4
19	2	22	5
		25	8

Quarto passo: Calcule as somas dos postos do Grupo 1 e do Grupo 2.

$$\sum R_1 = 3 + 1 + 6 + 2 = 12$$

$$\sum R_2 = 7 + 9 + 4 + + 5 + 8 = 33$$

Verifique os cálculos usando a fórmula (3.1):

$$\sum R_1 + \sum R_2 = n \times \frac{n+1}{2} = 12 + 33 = 9 \times \frac{9+1}{2} = 45$$

Quinto passo: A amostra é *muito pequena*. A distribuição da estatística *U não* se aproxima da normal padronizada – embora o resultado com essa aproximação apareça nas saídas de programas de computador. Procure, então, na Tabela 3 do Apêndice os valores críticos para $\sum R_1$. Para um teste unilateral com α = 0,05, você deve olhar na coluna de 0,05. Lembre que $n_1 = 4$ e $n_2 = 5$. Os valores críticos são 12-28. O valor calculado de $\sum R_1$ é 12. A diferença entre grupos é, portanto, significante no nível de 5%.

Sexto passo: A média dos postos para o Grupo 1 (homens) é 3; a média dos postos para o Grupo 2 (mulheres) é 8,25. A temperatura ambiente que homens consideram confortável é menor que a temperatura ambiente que mulheres consideram confortável. Essa diferença é significante no nível de 5%.

Se você usar o programa *Statistica*[35] vai obter resultados que precisam ser bem examinados (toda saída de computador deve ser estudada antes de se chegar a uma conclusão). O valor de *z* (não ajustado para empates) é igual ao valor de *z* ajustado porque não houve empates. O *p*-valor é não significante para um teste bilateral. No entanto, a proposta é proceder a um teste unilateral. Então, *p*-valor é 0,031746, portanto significante. No entanto, o valor calculado de *z* não tem distribuição que possa ser considerada aproximadamente normal padronizada.

Mann-Whitney U Test
By variable SEXO
Group 1: H Group 2: M

	Rank Sum	Rank Sum				Z		Valid N	Valid N	2*1sided
	Group H	Group M	U	Z	p-level	adjusted	p-level	Group H	GroupM	exact p
TEMPERAT	12	33	2	-1,96	0,05005	-1,9596	0,05005	4	5	0,06349

Você pode – e deve – usar um programa de computador para fazer os cálculos. No entanto, os pesquisadores precisam saber interpretar os resultados. Nas saídas dos programas podem aparecer diversas estatísticas – e é bom saber que nem todas devem ser usadas –, por isso é preciso fazer opções, mas com conhecimento de causa. Os programas em geral fornecem, no caso do teste de Mann-Whitney, as somas dos postos, as estatísticas de teste com e sem correção para empates com os respectivos p-valores e – às vezes – o p-valor obtido da normal padronizada, mesmo para amostras pequenas. Fique atento.

3.2.3. Teste da mediana

Indicação: O teste da mediana deve ser aplicado para testar a hipótese de que dois grupos independentes provieram de populações com a mesma mediana. É particularmente indicado quando existem dados censurados. Tem menor poder do que o teste de Mann-Whitney.

Veja como se faz o teste:

Primeiro passo: Estabeleça as hipóteses e o nível de significância.

Segundo passo: Junte os grupos, com n_1 e n_2 dados, em um só conjunto de $n_1 + n_2 = n$ dados. Calcule a *mediana* desse conjunto único de dados.

Terceiro passo: Conte o número de dados *iguais ou menores do que a mediana* e o número de dados *maiores do que a mediana* nos dois grupos. Arranje as contagens em tabela 2 × 2, como mostra o esquema.

Dados	**Grupo**	
	1	**2**
Menores ou iguais à mediana		
Maiores do que a mediana		

Quarto passo: Sob a hipótese de que os dois grupos provêm de populações com a mesma mediana, metade dos dados de cada grupo deve ser igual ou menor do que a mediana e metade deve ser maior do que a mediana. Aplique o teste de χ^2 para testar essa hipótese.[36]

Exemplo 3.7

Para testar a hipótese de que desencorajar pessoas afeta o rendimento delas em um teste de inteligência, 40 estudantes foram divididos ao acaso em dois grupos, o controle e o tratado.[37] Todos eles fizeram a primeira parte de um teste de inteligência. Duas semanas depois, os mesmos estudantes foram chamados para fazer a segunda parte do teste. O grupo controle fez essa parte do teste em condições normais, mas o grupo tratado foi desencorajado a fazer essa segunda parte porque – os pesquisadores informavam a eles – haviam obtido escores muito baixos na primeira parte. É preciso calcular as diferenças entre os escores obtidos na primeira e na segunda parte do teste para cada grupo de estudantes.

Diferenças entre os escores obtidos nas duas partes do teste de inteligência segundo o grupo

Grupo

Controle	Tratado
–1	7
8	–5
3	4
13	–4
0	–5

Diferenças entre os escores obtidos nas duas partes do teste de inteligência segundo o grupo (cont.)

Grupo	
Controle	**Tratado**
1	–7
6	–2
2	0
16	–6
3	6
14	–3
1	–3
4	–3
1	2
–3	–4
9	–3
3	1
3	–9
5	–4
2	0

Primeiro passo

- H_0: os grupos provêm de populações com a mesma mediana.
- H_1: os grupos provêm de populações com diferentes medianas.
- Nível de significância de 5%, teste bilateral.

Segundo passo: Combine os 20 + 20 = 40 dados em um só conjunto. A mediana é 1.

Terceiro passo: Conte, em cada grupo, quantos dados são menores ou iguais à mediana (em negrito na tabela auxiliar) e quantos são maiores.

Diferenças entre os escores obtidos nas duas partes do teste de inteligência segundo o grupo (Tabela auxiliar)

Grupo	
Controle	**Tratado**
–1	7
8	**–5**
3	4

13	**–4**
0	**–5**
1	**–7**
6	**–2**
2	**0**
16	**–6**

Diferenças entre os escores obtidos nas duas partes do teste de inteligência segundo o grupo (Tabela auxiliar) (cont.)

Grupo	
Controle	**Tratado**
3	6
14	**–3**
1	**–3**
4	**–3**
1	2
–3	**–4**
9	**–3**
3	**1**
3	**–9**
5	**–4**
2	0

Organize os resultados em uma tabela.

Distribuição das diferenças entre os escores obtidos nas duas partes do teste de inteligência segundo o grupo e seu valor em relação à mediana

Dados	**Grupo**		**Total**
	Controle	**Tratado**	
Menores ou iguais à mediana	6	16	22
Maiores do que a mediana	14	4	18
Total	20	20	40

Quarto passo: Aplique o teste de $\chi 2$. Neste exemplo, $\chi 2$ = 10,101, *p*-valor = 0,001482. Portanto, desencorajar pessoas invocando pretenso mau desempenho afeta o rendimento delas.

3.3. COMPARAÇÃO DE DOIS GRUPOS DEPENDENTES

3.3.1. Grupos dependentes

Dois grupos de dados são dependentes se cada unidade de um dos grupos corresponde a uma unidade do outro grupo. Uma dica para saber se dois grupos são dependentes é buscar, na descrição da pesquisa, palavras como dados pareados, dados emparelhados, antes e depois.

Exemplo 3.8

Para testar a eficácia de uma nova droga que, se presume, reduz a pressão arterial, uma indústria farmacêutica recrutou diversos voluntários que tinham pressão arterial alta. Foi, então, medida a pressão arterial de cada pessoa e, em seguida, fornecida a droga. Depois de a droga ter sido administrada, a pressão arterial dos voluntários foi, novamente, medida. Os dados são *dependentes* porque foram obtidos na mesma pessoa. Veja a Figura 3.2.

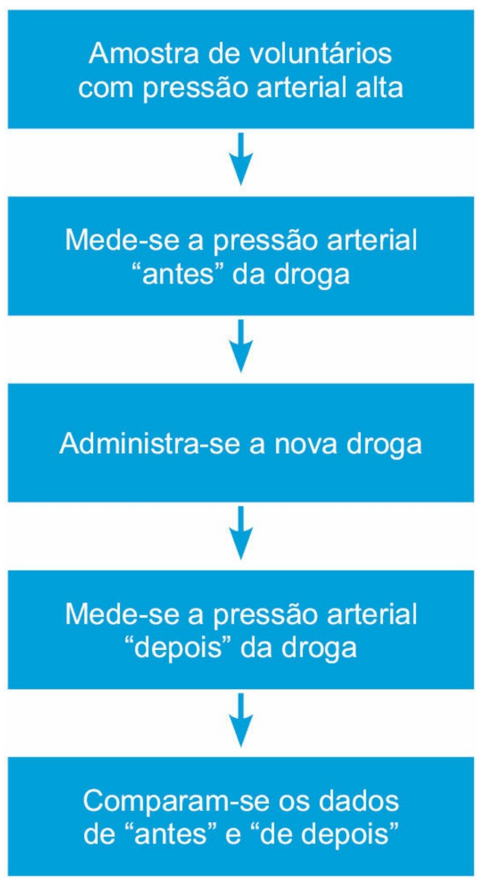


Figura 3.2 – Dados dependentes.

3.3.2. Teste dos postos assinalados de Wilcoxon

Indicação: O teste dos postos assinalados de Wilcoxon compara *dados pareados*. A variável em

análise deve ser *ordinal ou numérica*, porque o procedimento do teste exige calcular as diferenças entre os pares de dados. Veja como se faz o teste:

Primeiro passo: Estabeleça as hipóteses e o nível de significância.

Segundo passo: Calcule a diferença entre cada par de dados. Exclua toda diferença igual a zero.

Terceiro passo: Obtenha os *valores absolutos* das diferenças.

Quarto passo: Atribua postos aos *valores absolutos* das diferenças.

Quinto passo: Coloque *sinais nos postos*, que passam a se chamar *postos assinalados.* No entanto, observe os critérios:

- Postos das diferenças com *sinal negativo* ficam com *sinal negativo*.
- Postos das diferenças com *sinal positivo* ficam com *sinal positivo*.

Sexto passo: Some os postos assinalados. Indique a soma por SR. Some os quadrados dos postos assinalados. Indique a soma dos quadrados dos postos assinalados por $\sum R^2$.

Sétimo passo: Calcule a estatística de teste *z* que, sob a hipótese da nulidade, tem distribuição aproximadamente normal padronizada desde que $n \geq 20$.

$$z = \frac{\sum R}{\sqrt{\sum R^2}} \tag{3.5}$$

Oitavo passo: Faça o teste que consiste em comparar o valor calculado de *z* com o valor crítico dado na tabela de distribuição normal padronizada (Tabela 1 do Apêndice) no nível estabelecido de significância. Se você estiver usando um programa de computador, obterá o *p*-valor.

Exemplo 3.9

Foi feito um estudo para saber se a escuta é dicótica, isto é, se a capacidade de percepção auditiva tem grau diferente em cada um dos ouvidos.[38] Cada participante da pesquisa ouviu uma série de palavras, apresentadas aleatoriamente do lado direito e do lado esquerdo, e relatou o que ouviu. Ouve-se melhor de um lado do que de outro?

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa

Participante	**Lado**	
	Direito	**Esquerdo**
1	25	32
2	32	30
3	15	8
4	25	32
5	32	20
6	24	32
7	26	32
8	29	31

9	32	28
10	32	32
11	20	30
12	5	32
13	15	18
14	18	19
15	29	26
16	27	30
17	14	19
18	16	20
19	25	15
20	23	32

Para fazer o teste de Wilcoxon:

Primeiro passo: São n = 20 participantes de pesquisa.

- H_0: a capacidade de percepção auditiva tem o mesmo grau nos dois ouvidos. Teste bilateral, nível de significância de 5%.

Segundo passo: Calcule a diferença entre cada par de dados. Exclua toda diferença igual a zero.

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa (Tabela auxiliar)

Participante	**Lado**		**Diferença**
	Direito	**Esquerdo**	
1	25	32	–7
2	32	30	2
3	15	8	7
4	25	32	–7
5	32	20	12
6	24	32	–8
7	26	32	–6
8	29	31	–2
9	32	28	4
10	32	32	Excluído
11	20	30	–10
12	5	32	–27
13	15	18	–3
14	18	19	–1
15	19	26	–7
16	27	30	–3
17	14	19	–5
18	16	20	–4
19	25	15	10

20	23	32	–9

Terceiro passo: Obtenha os *valores absolutos* das diferenças.

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa (Tabela auxiliar)

Participante	Lado		Diferença	Valor absoluto
	Direito	Esquerdo		
1	25	32	–7	7
2	32	30	2	2
3	15	8	7	7
4	25	32	–7	7

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa (Tabela auxiliar) (cont.)

Participante	Lado		Diferença	Valor absoluto
	Direito	Esquerdo		
5	32	20	12	12
6	24	32	–8	8
7	26	32	–6	6
8	29	31	–2	2
9	32	28	4	4
10	32	32	Excluído	
11	20	30	–10	10
12	5	32	–27	27
13	15	18	–3	3
14	18	19	–1	1
15	19	26	–7	7
16	27	30	–3	3
17	14	19	–5	5
18	16	20	–4	4
19	25	15	10	10
20	23	32	–9	9

Quarto passo: Atribua postos aos *valores absolutos* das diferenças. Como são 19 participantes da pesquisa, isto é, n = 19 porque um foi excluído, os postos vão de 1 a 19. Verifique a igualdade definida pela fórmula (3.1).

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa (Tabela auxiliar)

Participante	Lado		Diferença	Valor absoluto	Posto
	Direito	Esquerdo			

1	25	32	–7	7	11,5
2	32	30	2	2	2,5
3	15	8	7	7	11,5
4	25	32	–7	7	11,5
5	32	20	12	12	18
6	24	32	–8	8	14
7	26	32	-6	6	9
8	29	31	-2	2	2,5
9	32	28	4	4	6,5
10	32	32	Excluído		
11	20	30	–10	10	16,5
12	5	32	–27	27	19
13	15	18	–3	3	4,5
14	18	19	–1	1	1

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa (Tabela auxiliar) (cont.)

Participante	**Lado**		**Diferença**	**Valor absoluto**	**Posto**
	Direito	**Esquerdo**			
15	19	26	–7	7	11,5
16	27	30	–3	3	4,5
17	14	19	–5	5	8
18	16	20	–4	4	6,5
19	25	15	10	10	16,5
20	23	32	–9	9	15
					$\sum R$ =190

$$\sum R = 19 \times \frac{19+1}{2} = 190$$

Quinto passo: Coloque *sinais nos postos*, que passam a se chamar *postos assinalados.* No entanto, observe os critérios:

- Postos de diferenças com *sinal negativo* ficam com *sinal negativo*.
- Postos de diferenças com *sinal positivo* ficam com *sinal positivo*.

Número de palavras ouvidas corretamente, segundo o lado em que foram ditas, para cada participante de pesquisa (Tabela auxiliar)

Participante	**Lado**		**Diferença**	**Valor absoluto**	**Posto**	**Posto assinalado**
	Direito	**Esquerdo**				
1	25	32	–7	7	11,5	–11,5
2	32	30	2	2	2,5	2,5
3	15	8	7	7	11,5	11,5

4	25	32	–7	7	11,5	–11,5
5	32	20	12	12	18	18
6	24	32	–8	8	14	–14
7	26	32	–6	6	9	–9
8	29	31	–2	2	2,5	–2,5
9	32	28	4	4	6,5	6,5
10	32	32	Excluído			
11	20	30	–10	10	16,5	–16,5
12	5	32	–27	27	19	–19
13	15	18	–3	3	4,5	–4,5
14	18	19	–1	1	1	–1
15	19	26	–7	7	11,5	–11,5
16	27	30	–3	3	4,5	–4,5
17	14	19	–5	5	8	–8
18	16	20	–4	4	6,5	–6,5
19	25	15	10	10	16,5	16,5
20	23	32	–9	9	15	–15

Sexto passo: Some os postos assinalados e some os quadrados dos postos assinalados.

A soma dos postos negativos é –135. A soma dos postos positivos é 55. Logo, ΣR = –80 e a soma dos quadrados dos postos assinalados é 2.463.

Sétimo passo: calcule a estatística de teste z que, sob a hipótese da nulidade, tem distribuição aproximadamente normal padronizada desde que[39] $n\ {}^{24}_{12}\ 25$.

$$z = \frac{-80}{\sqrt{2.463}} = -1{,}61$$

Oitavo passo: O valor crítico da distribuição normal padronizada (Tabela 1 do Apêndice) no nível de significância de 5% é ±1,96. Como o valor calculado de z = –1,61 é menor que o valor crítico, não se rejeita a hipótese de que a capacidade de percepção auditiva tenha o mesmo grau nos dois ouvidos. Um programa de computador dará p-valor = 0,1974.

Zeros e empates: Como *zero* não é nem positivo nem negativo, deve ser excluído, mas a amostra se reduz.

Um ou outro *empate* não muda o resultado do teste se a amostra for grande. No entanto, se houver muitos empates, a confiabilidade do teste fica reduzida. É preciso usar não a fórmula (3.6), mas uma fórmula com correção.[40]

Não é dada aqui a maneira de fazer essa correção porque aumenta, em muito, a dificuldade de cálculo. No entanto, convém saber que os programas para computador fazem a correção e fornecem o resultado do teste, com e sem a correção para empates. Uma opinião[41] interessante sobre esse assunto é a de que, se existem muitos zeros e muitos empates, o processo de medição deveria ser melhorado. No exemplo que vimos, a variável não é contínua – o que aumenta a

possibilidade de empates. O teste de Wilcoxon é mais confiável quando a variável é contínua. De qualquer forma, no exemplo, se a qualidade da medição melhorasse, o resultado seria mais confiável.

Amostras pequenas: Quando as amostras são pequenas, isto é, $n < 20$ e, particularmente, $n < 10$, a distribuição da variável z não se aproxima, satisfatoriamente, da distribuição normal padronizada. Nesses casos:

a) some os postos com sinal positivo;
b) some os postos com sinal negativo;
c) compare o valor absoluto das duas somas. A *menor soma de postos em valor absoluto é chamada de W*;
d) rejeite a hipótese da nulidade toda vez que o valor calculado de *W* for maior do que o valor dado na Tabela 4 do Apêndice, no nível estabelecido de significância.

Exemplo 3.10

Oito crianças participaram de um estudo sobre a efetividade de uma nova droga que se supõe reduzir o comportamento repetitivo de crianças autistas.[42] Foi medido o tempo em que cada criança se envolveu em comportamentos repetitivos durante um período de 3 horas tanto antes como depois de tomar o medicamento durante uma semana. Os dados estão apresentados a seguir.

Tempo despendido por crianças autistas em comportamentos repetitivos durante um período de observação de 3 horas segundo o grupo

Criança	**Grupo**	
	Antes da medicação	**Depois da medicação**
1	85	75
2	70	50
3	40	50
4	65	40
5	80	20
6	75	65
7	55	40
8	20	25

Primeiro passo:

- Hipótese da nulidade: a droga *não* tem efeito.
- Hipótese alternativa: a droga *tem* efeito.
- Nível de significância de 5%, teste bilateral.

Segundo passo: Organize os dados em uma tabela. Calcule as diferenças entre antes e depois de os pacientes terem sido medicados.

Tempo despendido por crianças autistas em comportamentos repetitivos durante um período de observação de 3 horas segundo o grupo (Tabela auxiliar – diferenças)

Criança	Grupo		Diferença entre grupos
	Antes da medicação	Depois da medicação	
1	85	75	10
2	70	50	20
3	40	50	–10
4	65	40	25
5	80	20	60
6	75	65	10
7	55	40	15
8	20	25	–5

Terceiro passo: Obtenha os *valores absolutos* das diferenças.

Tempo despendido por crianças autistas em comportamentos repetitivos durante um período de observação de 3 horas segundo o grupo (Tabela auxiliar – valores absolutos das diferenças)

Criança	Grupo		Diferença	Valor absoluto da diferença
	Antes	Depois		
1	85	75	10	10
2	70	50	20	20
3	40	50	–10	10
4	65	40	25	25
5	80	20	60	60
6	75	65	10	10
7	55	40	15	15
8	20	25	–5	5

Quarto passo: Atribua postos aos *valores absolutos* das diferenças. Como são oito os participantes da pesquisa, os postos vão de 1 a 8. Verifique a igualdade definida pela fórmula (3.1).

Tempo despendido por crianças autistas em comportamentos repetitivos durante um período de observação de 3 horas segundo o grupo (Tabela auxiliar – postos)

Criança	Grupo		Diferença	Valor absoluto da diferença	Posto
	Antes	Depois			
1	85	75	10	10	3
2	70	50	20	20	6
3	40	50	–10	10	3
4	65	40	25	25	7
5	80	20	60	60	8
6	75	65	10	10	3
7	55	40	15	15	5
8	20	25	–5	5	1

$$\sum R = 8 \times \frac{8+1}{2} = 36$$

Quinto passo: Coloque *sinais nos postos*, que passam a se chamar *postos assinalados.* Depois, some os postos com *sinal negativo* e com *sinal positivo.*

Tempo despendido por crianças autistas em comportamentos repetitivos durante um período de observação de 3 horas segundo o grupo (Tabela auxiliar – postos assinalados)

Criança	**Grupo**		**Diferença**	**Valor absoluto da diferença**	**Posto**	**Posto assinalado**
	Antes	**Depois**				
1	85	75	10	10	3	3
2	70	50	20	20	6	6
3	40	50	–10	10	3	–3
4	65	40	25	25	7	7
5	80	20	60	60	8	8
6	75	65	10	10	3	3
7	55	40	15	15	5	5
8	20	25	–5	5	1	–1

- Soma dos postos com sinal negativo: – 4.
- Soma dos postos com sinal positivo: +32.
- $\sum Ri$ = – 4 +32 = 28
- W = 4

Como a amostra é pequena (n = 8), procure o valor crítico para W na Tabela 4 do Apêndice (lembre-se de que W é a menor soma de postos em valor absoluto). Para α = 0,05 e amostras de tamanho n = 8, o valor crítico é 4. Como a menor soma obtida desprezando o sinal é W = 4, você pode afirmar, no nível de 5% de significância, que a nova droga tem efeito sobre o comportamento de crianças autistas.

3.3.3. Teste do sinal

Indicação: O teste do sinal é a forma mais fácil de comparar dois grupos dependentes. Mesmo assim, não é muito indicado porque tem pouco poder; usa como informação apenas o sinal das diferenças entre pares de dados.[43]

A única pressuposição necessária para a aplicação do teste do sinal é que a distribuição da variável seja contínua, porque usa o sinal das diferenças. Ainda, para usar a aproximação normal como mostrado aqui, é preciso ter grandes amostras.[44]

Veja como se faz o teste.

Primeiro passo: Estabeleça as hipóteses e o nível de significância.

Segundo passo: Compare o valor da primeira medida com o valor da segunda medida, feitas no mesmo par de pessoas, animais ou objetos; atribua sinal + (mais) para todo par de observações em que a primeira medida foi maior do que a segunda e sinal – (menos) quando acontecer o contrário.

Terceiro passo: Conte o número de sinais + (mais) e o número de sinais – (menos). Exclua toda diferença igual a zero.

Quarto passo: Sob a hipótese da nulidade, medidas feitas no mesmo par são iguais. Então, sob essa hipótese, metade (50%) das diferenças entre pares teria sinal + (mais) e a outra metade (50%) sinal – (menos). Logo, sob a hipótese da nulidade, a proporção de sinais – (menos) é 0,5. Para testar essa hipótese, calcule a estatística:

$$z = \frac{(|p - 0,5|) - \frac{1}{2 \times n}}{\sigma_p} \tag{3.7}$$

em que *n* é o tamanho da amostra descontados os pares de dados com diferença zero, *p* é a proporção de sinais – (menos) e σ_p é o desvio padrão[45] sob a hipótese da nulidade, isto é:

$$\sigma_p = \sqrt{\frac{0,5(1 - 0,5)}{n}} \tag{3.8}$$

Empates: Quando ocorre empate, a diferença fica igual a zero. Se o número de zeros for pequeno, eles podem ser ignorados como foi feito aqui, mas isso reduz o tamanho da amostra. Se o número de zeros for grande, adote outro procedimento.[46]

Quinto passo: Para fazer o teste, compare o valor calculado de *z* com o valor crítico dado na tabela de distribuição normal padronizada (Tabela 1 do Apêndice), no nível estabelecido de significância. Rejeite a hipótese da nulidade toda vez que o valor calculado de z for igual ou maior do que o valor de z dado na tabela. Se usar um programa de computador, achará o *p*-valor.

Exemplo 3.11

Para decidir pela compra de um de dois tomógrafos fabricados por empresas diferentes, o diretor de um instituto separou 20 crânios e fez duas tomadas tomográficas de cada um: a primeira usando o tomógrafo da marca A e a segunda usando o tomógrafo da marca B. Depois, pediu a um técnico que examinasse as tomografias e conferisse, a cada uma delas, uma nota de 0 (péssima) a 5 (excelente). Os resultados são dados em seguida.

Notas conferidas por um técnico para tomografias obtidas por dois tomógrafos

Crânio	Tomógrafo		Sinal da diferença
	A	B	
1	5	3	+

2	5	2	+
3	5	1	+
4	4	2	+
5	4	5	–
6	5	2	+
7	2	1	+
8	4	5	–
9	4	2	+
10	5	4	+
11	4	3	+
12	0	1	–
13	5	4	+
14	4	2	+
15	5	3	+
16	5	2	+
17	5	1	+
18	4	2	+
19	4	5	–
20	2	5	–

Primeiro passo: A hipótese da nulidade é a de que as tomografias feitas pelos tomógrafos das duas marcas têm a mesma qualidade. Seja α = 0,05, teste bilateral.

Segundo passo: Atribua sinal + (mais) para todo par de observações em que o tomógrafo da marca A recebeu nota mais alta do que o da marca B e sinal – (menos) quando aconteceu o contrário. Exclua toda diferença igual a zero.

Terceiro passo: O tomógrafo da marca A teve melhor avaliação 15 vezes e o da marca B, cinco vezes.

Quarto passo: Para aplicar a fórmula (3.7), você precisa obter a *proporção p* de sinais – (menos) na amostra e o desvio padrão σ_p, dado pela fórmula (3.8). Então:

$$p = \frac{5}{20} = 0{,}25$$

$$\sigma_p = \sqrt{\frac{0{,}5 \times (1 - 0{,}5)}{20}} = 0{,}11180$$

Para obter a estatística *z*, calcule:

$$z = \frac{(|0{,}25 - 0{,}5|) - \frac{1}{2 \times 20}}{0{,}11180} = 2{,}012$$

O valor crítico de *z* para $\alpha = 0{,}05$ é 1,96. Como o valor calculado *z* = 2,012 é maior do que o valor crítico, pode-se afirmar que o tomógrafo da marca A produz tomografias de melhor qualidade do que o da marca B.

Se você usar o programa *Statistica* para fazer os cálculos, obterá:

	Sign Test			
	No. of Non–ties	**Percent v < V**	**Z**	**p-level**
Tomo A & Tomo B	20	25	2,012461	0,044171

RESUMO E OBJETIVO DO CAPÍTULO

Lido este Capítulo, você será capaz não só de aplicar os testes apresentados, mas também de saber *por que* foram aplicados. São duas situações em que se recomenda o uso de estatística não paramétrica:

1. *Comparação de amostras independentes*: teste de Mann-Whitney ou teste da mediana.
2. *Comparação de duas amostras dependentes*: teste dos postos assinalados de Wilcoxon ou teste do sinal.

---

[28] Posto é a tradução para a palavra inglesa *rank*. Daí, o uso da letra *R* para indicar posto. Alguns estatísticos traduzem *rank* por ranque e usam a palavra ranquear para dizer conferir postos.

[29] Em inglês, *tie*.

[30] Se mais de um terço dos dados estiver envolvido em tais empates, é preciso buscar outra solução. Veja

LEHMANN, E.L. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden Day, 1975.CONOVER, W.J. *Practical Nonparametric Statistics.* 2nd ed. Nova York: Wiley, 1980.GIBBONS, J. D. *Nonparametric statistics: an introduction.* Newbury Park: Sage Publications, 1993.

[31] Em termos mais técnicos, o teste de Mann-Whitney é indicado para testar a hipótese de que as amostras provêm de populações tais que a probabilidade de uma observação, tomada ao acaso de uma população, ser maior do que outra observação tomada ao acaso da outra população é de 0,5.

[32] Para estudar distribuição normal, veja: VIEIRA, S. *Introdução à Bioestatística*. 5. ed. Rio de Janeiro: Elsevier, 2016.

[33] Mann Whitney U Test (Wilcoxon Rank Sum Test) Disponível em: http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric4.html. Acesso em 27 de outubro de 2017.

34 Veja:

LEHMANN, E.L. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden Day, 1975.

CONOVER, W.J. *Practical Nonparametric Statistics.* 2nd ed. Nova York: Wiley, 1980.

GIBBONS, J. D. *Nonparametric statistics: an introduction.* Newbury Park: Sage Publications, 1993.

HOLLANDER, M.; WOLFE, D.A. *Nonparametric Statistics Methods*. Nova York: Wiley, 1973.

35 6 Statsoft, Inc. 2300 East 14 th Street, Tulsa, OK 74104, USA.

[36] Veja o teste de χ^2 no Capítulo 5 deste livro.

[37] GORDON; DUREA. The effect of discouragement on the revised Stanford-Binet scale. *J Genetic Psychol* 73: 201-7, 1948. Apud LEHMANN, EL. *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden Day, 1975. p. 47. O exemplo é antigo, mas é interessante e é citado por Lehman, um clássico.

[38] The Wilcoxon test: users.sussex.ac.uk/~grahamh/.../WilcoxonHandoout2011.pdf. *Disponível em:* https://www.google.com.br/. Acesso *em 31 de outubro de 2017*.

[39] Não é o caso deste exemplo.

[40] Veja:

LEHMANN, EL. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden Day, 1975.

CONOVER, WJ. *Practical Nonparametric Statistics.* 2nd ed. Nova York: Wiley, 1980.

[41] GIBBONS, JD. *Nonparametric statistics: an introduction.* Newbury Park: Sage Publications, 1993. p. 18.

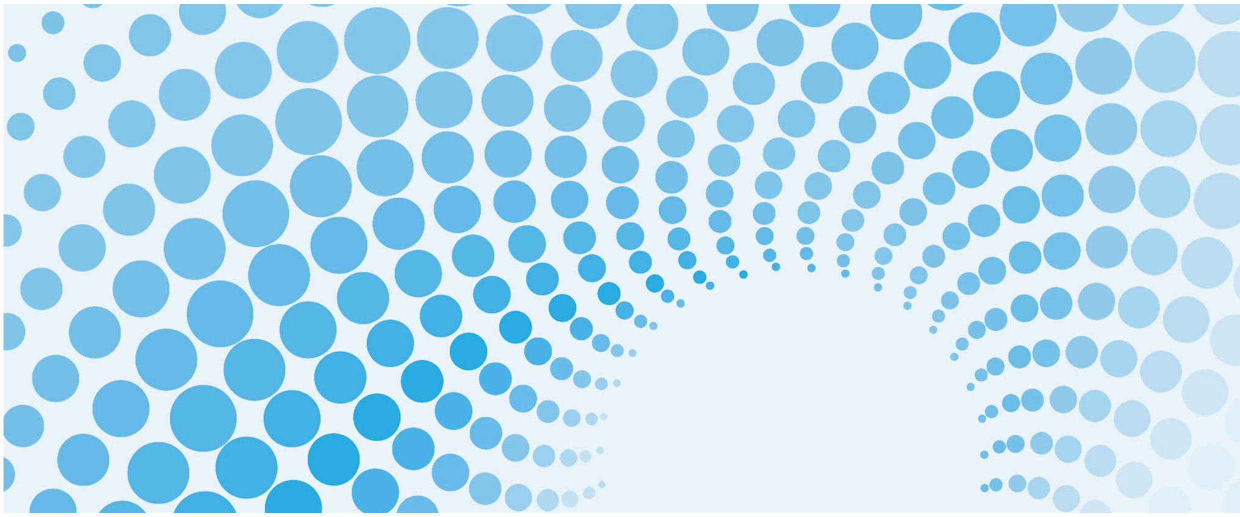
[42] Wilcoxon Rank Signed Test. Disponível em: http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704. Acesso em 1 de novembro de 2017.

[43] O teste *t* usa os valores das diferenças, o teste de Wilcoxon usa os postos das diferenças e o teste do sinal usa os sinais das diferenças.

[44] O teste do sinal usa a aproximação normal da binomial e a distribuição binomial aproxima-se da distribuição normal quando *n*┐ > 5 e *n* (1-┐) > 5.

[45] Desvio padrão na distribuição binomial.

[46] Para outras maneiras de proceder, veja, por exemplo: CONOVER, WJ. *Practical Nonparametric Statistics.* 2nd ed. Nova York: Wiley, 1980.



3.4.1. Considere o conjunto de dados: 86; 54; 57; 98; 57; 89; 43; 56; 67; 60. Atribua um posto a cada dado e verifique:

$$\sum R = n \times \frac{n+1}{2}$$

3.4.2. Imagine que um pesquisador quer determinar se duas linhagens de ratos de laboratório diferem na habilidade de correr em labirinto.[47] Para isso, seleciona uma amostra de ratos ingênuos (que nunca foram ensinados a correr em labirinto) de cada linhagem e dá o necessário treinamento. Faz, então, um teste: coloca os ratos para correr em labirinto e conta o número de tentativas de cada um até conseguir correr no labirinto duas vezes. Veja:

Linhagem A: 9; 4; 20; 13.

Linhagem B: 4; 12; 8; 3; 9.

a) Estabeleça a hipótese da nulidade e a hipótese alternativa. b) Aplique o teste de Mann-Whitney no nível de 5% de significância. Dê o valor de $\sum R1$ e os valores críticos de $\sum R1$. c) Conclua. d) Calcule $\sum R2$ e verifique que:

$$\sum R_1 + \sum R_2 = n \times \frac{n+1}{2}$$

3.4.3. Em uma pesquisa para estudar o efeito da inalação prolongada de cádmio sobre os níveis de hemoglobina, 10 cães foram expostos ao óxido de cádmio e 10 cães serviram como controle (não foram expostos ao óxido de cádmio). Os dados obtidos estão apresentados na Tabela 3.3. Faça o teste de Mann-Whitney.

Tabela 3.3 – Níveis de hemoglobina, em gramas, de cães segundo o grupo

Grupo	
Exposto	**Controle**
14,6	15,5
15,8	17,9
16,4	15,5
14,6	16,7
14,9	17,6
14,3	16,8
14,7	16,7
17,2	16,8
16,8	17,2
16,1	18,0

Fonte: DANIEL, W.W. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000. p.130.

3.4.4. Para testar o efeito de um novo analgésico nos casos de cefaleia, foi feito um ensaio clínico randomizado duplo-cego placebo-controlado com 17 pacientes. Oito pacientes receberam o novo analgésico (grupo tratado) e nove pacientes receberam placebo (grupo controle). Uma hora depois de ingerir o comprimido, os pacientes registraram a dor segundo uma escala visual analógica que variava de 0 a 10. Os dados estão na Tabela 3.4. Faça o teste de Mann-Whitney.

Tabela 3.4 – Registro de dor, medida em escala analógica, segundo o grupo

Grupo	
Tratado	**Controle**
1	2
1,5	3,5
2	4
2	5
3,5	8
5,5	8,5
7	9
7,5	9,5
	10

3.4.5. Dez restaurantes de uma cidade grande foram selecionados ao acaso para participar de um experimento conduzido pela Secretaria da Saúde como parte de um esforço para melhorar as condições de higiene nesse tipo de estabelecimento. Os chefes de cozinha foram pagos para assistir a um seminário de 3 dias no qual se enfatizaram as vantagens de uma cozinha limpa. As cozinhas dos restaurantes foram inspecionadas antes e 6 meses depois de os chefes terem assistido ao seminário. Nas duas ocasiões foram atribuídas as notas às condições de higiene das cozinhas, que estão na Tabela 3.5 . Faça o teste de Wilcoxon.

Tabela 3.5 – Notas das cozinhas dos restaurantes na inspeção antes e 6 meses depois de os chefes terem assistido ao seminário

Chefe de cozinha	Grupo	
	Antes	**Depois**
1	80	90
2	83	85
3	82	87
4	81	78
5	77	75
6	77	82
7	65	75
8	67	85
9	75	90
10	85	95

Fonte: DANIEL, W.W. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000. p.173.

3.4.6. Estudando o conformismo social e a autoestima, pesquisadores verificaram, aplicando o teste *t* de Student a uma amostra de 200 estudantes de pós-graduação, que o escore médio de autoestima no grupo caracterizado como conformista era diferente do escore médio de autoestima do grupo não conformista. Imagine que o estudo foi repetido com outros participantes e os pesquisadores chegaram aos dados apresentados na Tabela 3.6 . Faça o teste da mediana, porque esta amostra é muito menor do que a citada, de 200 estudantes.

Tabela 3.6 – Escore da autoestima segundo o grupo

Grupo	
Conformista	**Não conformista**

48	59
55	58
56	48
49	57
41	59
55	45
44	59
53	68
42	61
50	

Fonte: DANIEL, W.W. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000. p.131.

3.4.7. Foi feito um ensaio clínico com oito pacientes que iriam se submeter à extração de vários dentes para testar o efeito de um ansiolítico. Os níveis de ansiedade dos pacientes, antes e depois de medicados, registrados usando escala apropriada, estão na Tabela 3.7.

Tabela 3.7 – Níveis de ansiedade antes e depois de os pacientes terem sido medicados

Paciente	**Grupo**	
	Antes da medicação	**Depois da medicação**
1	23	14
2	26	14
3	28	29
4	29	25
5	37	31
6	26	18
7	27	30
8	32	30

3.4.8. Foi feito um ensaio para estudar os efeitos adversos de um fármaco indicado para casos de hipertensão. Os pacientes, com o mesmo perfil, recém-diagnosticados e sem tratamento, responderam a diversas questões ao iniciarem o tratamento e na oitava semana após o início. Os dados relativos à fadiga estão apresentados na Tabela 3.8. Neste exemplo, 0 significa ausência do sintoma; 1: o sintoma existe, mas não incomoda; 2: o sintoma incomoda, mas não interfere nas atividades; 3: interfere na maioria das atividades. Faça o teste do sinal.

Tabela 3.8 – Fadiga, medida em uma escala de quatro pontos segundo o grupo

Paciente	**Grupo**	
	Início	**Final**
1	1	2
2	2	3
3	0	2
4	3	0
5	1	0

6	3	1
7	3	2
8	3	3
9	2	2
10	2	0
11	3	1
12	0	0
13	2	2
14	2	3
15	2	1
16	3	0
17	2	0
18	3	1
19	3	0
20	2	2
21	2	3
22	3	1
23	3	0

Fonte: Dados fictícios com base em LUNA, RL. Eficácia e tolerabilidade da associação bisoprolol/hidroclorotiazida na hipertensão arterial. Arq Bras Cardiol 71(4), 1998.

3.4.9. Foram comparados os tempos de resolução nos casos de óbito intrauterino ocorridos no segundo trimestre de gestação para gestantes submetidas ao misoprostol administrado por vias diferentes. Os dados estão apresentados na Tabela 3.8. Faça o teste de Mann-Whitney.

Tabela 3.9 – Tempo de resolução, em horas, do óbito fetal ocorrido no segundo trimestre de gestação: comparação entre as vias oral e vaginal

Grupo	
Oral	**Vaginal**
30	24
9	11
22	20
10	22
20	27
16	6
	10
	23

Fonte: MACAU, S.N.; BERTINI, M.; CAMANO, L. Resolução do óbito intrauterino com o uso de misoprostol intravaginal em baixa dose. *Rev Bras Ginecol Obstet* 5: 232-236, 1992.

3.4.10. Uma pesquisa comparou a efetividade da propaganda feita por duas marcas rivais de café, A e B. Pessoas que estavam em um *shopping center* receberam, aleatoriamente, ou a propaganda feita por A ou a propaganda feita por B e disseram, entre

0 e 10, a probabilidade de comprar café da marca exibida na propaganda que lhes era mostrada. Os dados estão na Tabela 5.10. Faça o teste de Mann-Whitney (grupos independentes).[48] A propaganda de uma das marcas se revelou mais efetiva?

Tabela 3.10 – Notas conferidas a duas marcas de café por dois grupos de pessoas

Marca	
A	**B**
3	9
4	7
2	5
6	10
2	6
5	8

[47] O exercício é de MINIUM, E.W.; CLARKE, R.C.; COLADARCI, T. *Elements of Statistical Reasoning.* 2nd ed. New York: Wiley, 1999. p.421.

[48] Veja a solução detalhada em:

Mann-Whitney test worked example. Disponível em: http://users.sussex.ac.uk/~grahamh/RM1web/Mann-Whitney%20worked%20example.pdf. Acesso em 4 de novembro de 2017.

4

Testes não Paramétricos para Comparar mais de Dois Grupos

Na área da saúde, há uma busca incessante por novas formas de prevenir, detectar ou tratar doenças. Também se procura uma maneira de melhorar a qualidade de vida das pessoas com doenças crônicas ou minorar o sofrimento no final da vida. No decorrer de todas as fases de tais pesquisas são aplicados conhecimentos de estatística, seja para determinar o tamanho da amostra, delinear um ensaio ou fazer a análise dos dados coletados.

Em termos de análise estatística – que é o assunto deste livro –, em linhas gerais pode-se propor o que se segue.

Para grupos independentes:

- Se a variável for numérica e tiver distribuição normal ou aproximadamente normal, ou tiver apenas distribuição simétrica mas a amostra for bastante grande, faça uma análise de variância.[49]
- Se a variável for ordinal ou, se numérica, mas a distribuição não for simétrica e/ou a amostra for pequena, aplique um teste não paramétrico: o teste de Kruskal-Wallis ou o teste da mediana, mostrados neste Capítulo.

Para grupos dependentes:

- Se a variável for numérica e tiver distribuição normal ou aproximadamente normal, ou tiver apenas distribuição simétrica mas a amostra for bastante grande, faça uma análise de variância.[50]
- Se a variável for ordinal ou numérica, mas a distribuição não for simétrica e/ou a amostra for pequena, aplique um teste não paramétrico: o teste de Friedman, que veremos neste Capítulo.

4.1. COMPARAÇÃO DE MAIS DE DOIS GRUPOS INDEPENDENTES

4.1.1. Teste de Kruskal-Wallis

O teste de Kruskal-Wallis também é conhecido como análise de variância de Kruskal-Wallis, ANOVA de Kruskal-Wallis, ANOVA não paramétrica.

Indicação: O teste de Kruskal-Wallis é indicado para comparar três grupos, como magros, saudáveis, obesos, ou mais de três grupos, como mês da observação (janeiro, fevereiro, março etc.). A variável em análise deve ser ordinal (como escala visual analógica ou itens de uma escala Likert) ou contínua (como peso, tempo, QI). As observações devem ser independentes, isto é, as unidades de um grupo não devem estar relacionadas com as unidades do outro grupo, nem devem estar relacionadas entre si. A hipótese em teste é a de que postos dos grupos em comparação têm a mesma distribuição.

Importante: O teste de Kruskal-Wallis trabalha com *postos* – não com dados coletados. Portanto, *não* testa a hipótese de igualdade de médias da variável em estudo.

A lógica do teste de Kruskal-Wallis é muito semelhante à do teste de Mann-Whitney. Se você observou k grupos, indique o número de casos em cada grupo por n_1, n_2,... n_k. Então, $n_1 + n_2 +... + n_k = n$. Para fazer o teste de Kruskal-Wallis:

Primeiro passo: Estabeleça a hipótese da nulidade, a hipótese alternativa e o nível de significância.

Segundo passo: Junte os $n_1 + n_2 +... + n_k = n$ dados em um só conjunto. Atribua um posto a cada dado, *sem perder* a identificação do grupo ao qual cada dado pertence. Os postos vão, portanto, de 1 (o dado de menor valor) até n (o dado de maior valor). Os valores empatados recebem postos correspondentes à média dos dos postos que lhes seriam conferidos.

Terceiro passo: Calcule as somas dos postos das unidades de cada grupo, isto é, õR_1, õR_2..., õR_k. Se os cálculos estiverem corretos, tem-se que:

$$\sum R_1 + \sum R_2 + \ldots + \sum R_k = n \times \frac{n+1}{2} \tag{4.1}$$

Quarto passo: Se não houver empates (ou os empates forem poucos), calcule:

$$H = \frac{12}{n(n+1)}\left[\frac{\left(\sum R_1\right)^2}{n_1} + \frac{\left(\sum R_2\right)^2}{n_2} + \ldots + \frac{\left(\sum R_k\right)^2}{n_k}\right] - 3(n+1) \tag{4.2}$$

em que ΣR_1, ΣR_2,..., ΣR_k são as somas dos postos das unidades nos grupos 1, 2,..., k, respectivamente; n_1, n_2,..., n_k representam o número de unidades nos respectivos grupos; n é o número total de unidades, isto é, $n_1 + n_2 + ... + n_k$.

Quinto passo: Se a amostra for grande e houver vários grupos em comparação, sob a hipótese da nulidade a estatística H tem, aproximadamente, distribuição de $\chi 2$ com (k – 1) graus de liberdade. Compare, então, o valor calculado de H com o valor crítico dado na tabela de $\chi 2$ (Tabela 4 do Apêndice) ao nível estabelecido de significância e com (k – 1) graus de liberdade. Rejeite a hipótese da nulidade toda vez que o valor calculado de H for igual ou maior do que o valor encontrado na tabela. Se estiver usando um programa para computador, procure o p-valor associado à estatística H.

Sexto passo: Para facilitar a comparação, calcule as *medianas* ou as *médias dos postos* dos grupos.

Exemplo 4.1

Para avaliar o efeito da expectativa sobre a percepção da qualidade na degustação de bebidas, um pesquisador convidou 24 apreciadores de vinho para uma sessão de degustação e os dividiu aleatoriamente em três grupos de mesmo tamanho.[51] Um dos participantes do grupo B e dois do grupo C não compareceram. O tamanho da amostra ficou, portanto, reduzido a 21 participantes. Aqueles que compareceram foram convidados a avaliar a qualidade de determinado vinho em uma escala visual analógica que ia de 0 (péssimo) a 10 (excelente). Cada pessoa fazia uma entrevista individual. Para as pessoas do grupo A se induzia a expectativa de um vinho de ótima qualidade. Para as pessoas do grupo B se induzia a expectativa de um vinho de baixa qualidade e para as pessoas do grupo C não se faziam previsões. O vinho era o mesmo. A nota conferida pelos degustadores, segundo o grupo ao qual foram designados, é a variável que será posta em análise.

Notas conferidas pelos degustadores segundo o grupo

	Grupo	
A	**B**	**C**
6,4	2,5	1,3
6,8	3,7	4,1
7,2	4,9	4,9
8,3	5,4	5,2
8,4	5,9	5,5
9,1	8,1	8,2
9,4	8,2	
9,7		

Primeiro passo:

- Hipótese da nulidade: os postos dos grupos em comparação são estatisticamente iguais.
- Nível de significância: $\alpha = 0{,}05$ para um teste bilateral.

Segundo passo: Junte os 21 dados em um só conjunto. Atribua um posto a cada dado. O posto 1

deve ser atribuído para o valor mais baixo (que é 1,3) e o posto 21 para o valor mais alto (que é 9,7), mas não perca a identificação do grupo a que o dado pertence. Os valores empatados recebem postos correspondentes à média dos postos que lhes seriam atribuídos.

Notas conferidas segundo o grupo de degustadores (Tabela auxiliar: postos no conjunto)

Grupo	**Nota**	**Posto**
C	1,3	1
B	2,5	2
B	3,7	3
C	4,1	4
B	4,9	5,5
C	4,9	5,5
C	5,2	7
B	5,4	8
C	5,5	9
B	5,9	10
A	6,4	11
A	6,8	12
A	7,2	13
B	8,1	14
B	8,2	15,5
C	8,2	15,5
A	8,3	17
A	8,4	18
A	9,1	19
A	9,4	20
A	9,7	21

Terceiro passo: Para cada grupo, calcule as somas dos postos, isto é, ΣR_1, ΣR_2 e ΣR_3.

Notas conferidas segundo o grupo de degustadores (Tabela auxiliar)

Grupo	**Nota**	**Posto**	**Grupo**	**Nota**	**Posto**	**Grupo**	**Nota**	**Posto**
A	6,4	11	B	2,5	2	C	1,3	1
A	6,8	12	B	3,7	3	C	4,1	4
A	7,2	13	B	4,9	5,5	C	4,9	5,5
A	8,3	17	B	5,4	8	C	5,2	7
A	8,4	18	B	5,9	10	C	5,5	9
A	9,1	19	B	8,1	14	C	8,2	15,5
A	9,4	20	B	8,2	15,5			
A	9,7	21						

131	58	42

Para conferir os cálculos, lembre-se da fórmula (4.1). Então:

$$131 + 58 + 42 = 21 \times \frac{22}{2} = 231$$

Quarto passo: Calcule *H* conforme a fórmula (4.2):

$$H = \frac{12}{21(21+1))}\left[\frac{131^2}{8} + \frac{58^2}{7} + \frac{42^2}{6}\right] - 3(21+1)$$

$$H = \frac{12}{462}[2145{,}125 + 480{,}5714 + 294\ ] - 66 = 9{,}836$$

Quinto passo: Sob a hipótese da nulidade a estatística *H* tem, aproximadamente, distribuição de $\chi 2$ com (3 – 1)=2 graus de liberdade. O valor crítico de $\chi 2$ no nível de significância de 5% e com *2* graus de liberdade é 5,99. Rejeite a hipótese de que os postos médios dos três grupos em comparação são estatisticamente iguais. Se usar um programa para computador, obterá *p*-valor = 0,0073.

Sexto passo: Para facilitar a comparação, calcule as *médias dos postos* dos grupos. Obterá:

- Média dos postos do Grupo A = 16,375
- Média dos postos do Grupo B = 8,286
- Média dos postos do Grupo C = 7,000

Amostras pequenas

Se houver *apenas três grupos em comparação e cinco ou menos observações em cada grupo*, a distribuição da estatística *H* não se aproxima, satisfatoriamente, da distribuição de $\chi 2$. Nesses casos, você deve recorrer às tabelas que dão os valores críticos de *H* para alguns níveis de significância (Tabela 5 do Apêndice).

Exemplo 4.2

Foi feito um estudo para verificar se a idade é fator de risco para a gravidade da pneumonia hospitalar em pacientes internados em unidade de terapia intensiva (UTI). Foram obtidas as idades de 15 pacientes, classificados em três grupos segundo a gravidade do quadro clínico da pneumonia: moderada, grave, muito grave. Verifique se a idade do paciente é fator de risco para a gravidade da pneumonia hospitalar em pacientes na UTI.

Idade dos pacientes internados em UTI segundo a gravidade do quadro clínico da pneumonia hospitalar

	Grupo	
Moderada	**Grave**	**Muito grave**
54	69	87
26	48	84

43	51	71
39	36	83
58	40	67

Primeiro passo:

- Hipótese da nulidade: os postos médios dos três grupos em comparação são estatisticamente iguais.
- Nível de significância α = 0,05, teste bilateral.

Segundo passo: Junte os 5 + 5 + 5 = 15 dados em um só conjunto. Atribua posto 1 ao valor mais baixo (que é 26) e o posto 15 para o valor mais alto (que é 87), mas não perca a identificação do grupo a que o dado pertence. Note que não houve empates.

Terceiro passo: Para cada grupo calcule as somas dos postos, isto é, ΣR_1, ΣR_2 e ΣR_3. Para conferir os cálculos, lembre-se da fórmula (4.1):

$$26 + 30 + 64 = 15 \times \frac{15+1}{2} = 120$$

Idade dos pacientes internados em UTI segundo a gravidade do quadro clínico da pneumonia hospitalar (Tabela auxiliar: postos para o teste de Kruskal-Wallis)

Grupo					
Moderada		Grave		Muito grave	
Dado	Posto	Dado	Posto	Dado	Posto
54	8	69	11	87	15
26	1	48	6	84	14
43	5	51	7	71	12
39	3	36	2	83	13
58	9	40	4	67	10
	$\Sigma R_1 = 26$		$\Sigma R_2 = 30$		$\Sigma R_3 = 64$

Quarto passo: Calcule *H*, conforme a fórmula (4.2):

$$H = \frac{12}{15(15+1)}\left[\frac{26^2}{5} + \frac{30^2}{5} + \frac{64^2}{5}\right] - 3(15+1) = 8{,}720$$

Quinto passo: como a amostra é pequena (são três grupos de cinco observações), é melhor recorrer às tabelas que dão os valores críticos de *H* para alguns níveis de significância (Tabela 5 do Apêndice). Lá você encontra, para α = 0,05 e três grupos de cinco observações (linha em que se lê 5 5 5), o valor crítico 5,780. Rejeite, então, a hipótese de que a idade de pacientes com pneumonia hospitalar, internados em UTI, seja a mesma, qualquer que seja o nível de gravidade do quadro clínico.

Se você usar o programa *Statistica*[52] vai obter resultados que precisam ser bem examinados (toda saída de computador deve ser estudada antes de se chegar a uma conclusão). O valor de *H* (não ajustado para empates porque não houve empates) é 8,720, com 2 graus de liberdade. O *p*-valor indica significância para um teste bilateral.

Kruskal-Wallis ANOVA by Ranks

Independent (grouping) variable: idade			
Kruskal-Wallis test: H (2, N= 15) = 8,720001 p =,0128			
	Code	**Valid N**	**Sum of Ranks**
Group 1	1	5	26
Group 2	2	5	30
Group 3	3	5	64

Empates

Fique atento aos empates: se eles ocorreram, a fórmula (4.2) precisa de uma correção. Quanto *menor* for a amostra e quanto *maior* for o número de empates, mais necessária será a correção. Os cálculos são trabalhosos e os programas para computador fazem a correção automaticamente. De qualquer modo, veja como se faz a correção. É preciso calcular:

$$C = 1 - \frac{e_1\left(e_1^2 - 1\right) + e_2\left(e_2^2 - 1\right) + \ldots + e_m\left(e_m^2 - 1\right)}{n\left(n^2 + 1\right)} \tag{4.3}$$

Nesta fórmula, m é o número de grupos de empates; e_i (i = 1, 2,..., m) é o número de valores empatados no i-ésimo grupo. A estatística de teste para o teste de Kruskal-Wallis fica, então, como segue:

$$H^* = \frac{H}{C} \tag{4.4}$$

Na fórmula (4.4), H^* é a estatística, H é a estatística calculada em (4.2) e C é a correção calculada em (4.3).

Exemplo 4.3

Foi feito um ensaio clínico randomizado duplo-cego com pacientes com osteoartrite para comparar o efeito do etoricoxib e do naproxeno sobre a dor. Os pacientes foram randomizados para receber placebo (dois comprimidos ao dia) ou etoricoxib (60 mg uma vez ao dia) ou naproxeno (500 mg duas vezes ao dia). O estudo durou 12 semanas. Os próprios pacientes avaliaram, à noite, a dor que sentiam usando uma escala visual analógica (EVA) (0 = nenhuma dor; 100 = dor extrema). A diferença entre os registros de dor, feitos pelos pacientes no início e no final do ensaio (diminuição da dor), estão na tabela.

Diminuição da dor registrada na EVA considerando início e final do ensaio

	Grupo	
Placebo	**Etoricoxib**	**Naproxeno**
10	20	27
21	28	30
11	30	22

19	25	23
15	25	25
10	26	29
20	30	25

Fonte: Dados fictícios, baseados em LEUNG, A.T. et al. Efficacy and tolerability profile of etoricoxib in patients with osteoarthritis: a randomized, double-blind, placebo and active-comparator controlled 12week efficacy trial. *Curr Med Res Opin* 18(2):49-58, 2002.

Para fazer a correção no caso de empates, você já deve ter atribuído um posto a cada dado, começando com o posto 1 para o valor mais baixo (que é 10) e terminando com o posto 21 para o valor mais alto (que é 30), sem perder a identificação do grupo ao qual cada dado pertence. No caso dos empates, deve ser atribuída a média dos postos que eles receberiam na sequência. São $m = 4$ grupos de empates.

Diminuição da dor registrada na EVA, considerando início e final do ensaio (Tabela auxiliar: postos)

Grupo	**EVA**	**Posto**	**Grupo**	**EVA**	**Posto**	**Grupo**	**EVA**	**Posto**
Placebo	10	1,5	Placebo	21	8	Etoricoxib	26	15
Placebo	10	1,5	Naproxeno	22	9	Naproxeno	27	16
Placebo	11	3	Naproxeno	23	10	Etoricoxib	28	17
Placebo	15	4	Etoricoxib	25	12,5	Naproxeno	29	18
Placebo	19	5	Etoricoxib	25	12,5	Etoricoxib	30	20
Etoricoxib	20	6,5	Naproxeno	25	12,5	Etoricoxib	30	20
Placebo	20	6,5	Naproxeno	25	12,5	Naproxeno	30	20

1º grupo de empates: 10; 10. Logo $e_1 = 2$

2º grupo de empates: 20; 20. Logo $e_2 = 2$

3º grupo de empates: 25; 25; 25; 25. Logo $e_3 = 4$

4º grupo de empates: 30; 30; 30. Logo $e_4 = 3$

$$C = 1 - \frac{2\left(2^2-1\right)+2\left(2^2-1\right)+4\left(4^2-1\right)+3\left(2^2-1\right)}{21\left(21^2+1\right)} = 0{,}989657$$

Logo

$$H^* = \frac{H}{C} = \frac{12{,}6141}{0{,}989657} = 12{,}745926$$

Para o teste de Kruskal Wallis, sem correção para empates, $H = 12{,}6141$. *Com correção para os empates,* $H^* = 12{,}7459$. A diferença não é grande, mas faça a correção sempre que dispuser de um programa para isso.

Médias e medianas

Quando se aplica um teste de Kruskal-Wallis, não devem ser apontadas nem médias nem medianas dos dados, nem devem ser apresentados gráficos com essas estatísticas. Comparam-se médias e medianas dos *postos*. Para deixar isso claro, é apresentado um exemplo bastante

engenhoso[53] que compara três grupos com médias e medianas iguais, mas com postos médios diferentes. Logo, os grupos são diferentes.

Exemplo 4.4

O resultado do teste de Kruskal-Wallis para os dados apresentados em seguida é significante: $H = 7{,}355$, p-valor = 0,0253. No entanto, os três grupos têm a mesma média (43,5) e a mesma mediana (27,5). Então, apresentar as médias ou as medianas dos três grupos não mostra que as distribuições são diferentes. Os postos médios são diferentes: 34,6; 27,5 e 20,4, respectivamente.

	Grupo	
A	**B**	**C**
1	10	19
2	11	20
3	12	21
4	13	22
5	14	23
6	15	24
7	16	25
8	17	26
9	18	27
46	37	28
47	58	65
48	59	66
49	60	67
50	61	68
51	62	69
52	63	70
53	64	71
342	193	72

4.1.1.1.Teste de Dunn para comparações múltiplas

Os pesquisadores não ficam satisfeitos com a conclusão de que as populações amostradas não são iguais ou de que os tratamentos estudados não têm, todos, o mesmo efeito. Eles querem saber onde estão as diferenças. Nesses casos, é muito usado o teste de Dunn.[54]

Indicação: O teste de Dunn compara postos médios de grupos, dois a dois. *Não* exige que os grupos tenham o mesmo número de observações, mas tem melhor aproximação quando as amostras são grandes.

Para fazer o teste:

Primeiro passo:

Hipótese da nulidade: os grupos em comparação têm os mesmos postos médios.[55] Estabeleça o nível de significância.

Segundo passo: Calcule os postos médios dos k grupos.

$$\overline{R}_i = \frac{\sum R_i}{n_i}$$

em que $i = 1,2,...,k$ indica o posto médio do i-ésimo grupo.

Terceiro passo: Calcule o erro padrão para cada par de postos médios.

$$SE = \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \tag{4.5}$$

Quarto passo: Calcule as estatísticas de teste para comparar as médias de postos, duas a duas:

$$q = \frac{\overline{R}_i - \overline{R}_j}{SE} \tag{4.6}$$

em que i e j indicam o i-ésimo e o j-ésimo grupo, $i = 1, 2,...,k$, $j = 1, 2,...k$ e $i \neq j$. Toda vez que o valor calculado de q for maior do que o valor crítico de q (Tabela 6 do Apêndice) no nível estabelecido de significância e para k grupos, a diferença entre grupos é significante.[56] Se você estiver usando um programa para computador, procure o p-valor associado a cada diferença de postos médios.

Empates

Fique atento aos empates: se eles ocorreram, a fórmula (4.6) precisa de correção. Quanto *menor* for a amostra e quanto *maior* for o número de empates, mais necessária será a correção. Faça os cálculos usando um computador, uma vez que os programas já fazem a correção automaticamente.

Amostras pequenas

O teste de Dunn tem melhor aproximação quando as amostras são grandes. Quando há *apenas três grupos em comparação e cinco ou menos observações em cada grupo* – como acontece no exemplo –, recomenda-se fazer a correção.[57] Nem todo programa para computador faz essa correção específica. Então, procure um programa que a faça.

Exemplo 4.5

Reveja o Exemplo 4.2. Para comparar os grupos, dois a dois, aplique o teste de Dunn.

Primeiro passo:

• Hipótese da nulidade: os grupos em comparação têm os mesmos postos médios.[58]

• Hipótese alternativa: pelo menos um grupo tem posto médio diferente.

• Nível de significância ∼ Ť ŕĜŕṕh

Segundo passo: Calcule os postos médios dos três grupos:

$$\overline{R}_1 = \frac{26}{5} = 5{,}2$$

$$\overline{R}_2 = \frac{30}{5} = 6{,}0$$

$$\overline{R}_3 = \frac{64}{5} = 12{,}8$$

Terceiro passo: Calcule o erro padrão da diferença de cada dois postos médios, que é dado pela fórmula (4.5):

$$SE = \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Como, neste exemplo, todos os grupos têm o mesmo tamanho, o erro padrão da diferença de dois postos médios, para todas as comparações é:

$$SE = \sqrt{\frac{15(15+1)}{12}\left(\frac{1}{5} + \frac{1}{5}\right)} = \sqrt{8} = 2{,}8284$$

Quarto passo: Para comparar postos médios dois a dois, calcule as estatísticas de teste pela fórmula (4.6) e compare com o valor crítico. Veja a tabela resumo para o teste de Dunn: o posto médio do terceiro grupo é significativamente maior do que os outros dois. Então, pacientes mais velhos internados em UTI têm pneumonia hospitalar mais grave.

Tabela resumo para o teste de Dunn aos dados apresentados no exemplo 4.1

Comparação	$\overline{R}_i - \overline{R}_j$	***q***	***q* crítico**	**Conclusão**
Moderada versus grave	6 – 5,2 = 0,8	0,282843	2,394	Não rejeita H_0
Moderada versus muito grave	12,8 – 5,2 = 7,6	2,687006	2,394	Rejeita H_0
Grave versus muito grave	12,8 – 6,0 = 6,8	2,404163	2,394	Rejeita H_0

Se você usar o programa *SigmaStat* para fazer os cálculos, obterá:

All Pairwise Multiple Comparison Procedures (Dunn’s Method)

Comparison	**Diff of Ranks**	**Q**	**P<0,05**
Col 3 vs Col 1	7,600	2,687	Yes
Col 3 vs Col 2	6,800	2,404	Yes
Col 2 vs Col 1	0,800	0,283	No

Note: The multiple comparisons on ranks do not include an adjustment for ties.

4.1.2.Teste da mediana

Indicação: O teste da mediana é indicado para testar a hipótese de que três ou mais populações provieram de populações com a mesma mediana. A variável em análise deve ser ordinal ou numérica.

Importante: O teste da mediana é particularmente útil quando existem dados censurados (alguns dados estão aquém ou além dos limites estabelecidos para coleta).

Para fazer o teste da mediana:

Primeiro passo: Sejam k grupos com n_1, n_2,..., n_k *dados.* Estabeleça as hipóteses e o nível de significância.

Segundo passo: Junte os k grupos em comparação em um só conjunto. Depois, calcule a mediana geral desse conjunto de dados.

Terceiro passo: Conte, em cada grupo, o número de dados menores ou iguais à mediana e o número de dados maiores do que essa mediana em todos. Arranje as contagens em uma tabela $2 \times k$.

Número de dados	**Grupo**			
	1	**2**	**...**	k
Menores ou iguais à mediana				
Maiores do que a mediana				

Quarto passo: Sob a hipótese de que todos os grupos vieram de populações com a mesma mediana, metade dos dados de cada grupo deve ser igual ou menor do que a mediana e metade deve ser maior. Aplique o teste de $\chi 2$ para testar essa hipótese.[59]

Exemplo 4.6

Para saber se o nível de estresse é maior em fumantes, em não fumantes ou em pessoas que deixaram o hábito, um pesquisador avaliou 30 bancários usando uma escala validada. Foram retiradas da amostra todas as pessoas que, na ocasião, estivessem enfrentando situações muito estressantes. Como a variável é ordinal, você pode aplicar o teste da mediana.

Nível de estresse para fumantes,não fumantes e ex-fumantes

	Grupo	
Fumante	**Não fumante**	**Ex-fumante**
25	26	30
26	35	32
56	65	50
45	64	34
46	52	35

52	53	64
40	64	71
42	52	74
65	23	56
53	25	38

Primeiro passo: A hipótese em teste é a de que os grupos vieram de populações com a mesma mediana. Seja α = 0,05.

Segundo passo: Combine os 10 + 10 + 10 = 30 dados em um só conjunto. Encontre a mediana de todos os 30 dados, que é 48.

Terceiro passo: Conte, em cada grupo, quantos dados são menores ou iguais à mediana e quantos são maiores. Apresente esses dados em uma tabela auxiliar, o que facilitará a aplicação do teste.

Tabela auxiliar para proceder ao teste da mediana

Número de dados	**Grupo**		
	Fumante	**Não fumante**	**Ex-fumante**
Menores ou iguais à mediana	6	4	5
Maiores do que a mediana	4	6	5
Total	10	10	10

Quarto passo: Para testar a hipótese de que todos os grupos vieram de populações com a mesma mediana, aplique o teste de χ^2. Nesse exemplo, χ^2 = 0,800. No nível de significância α = 0,05 e com 2 graus de liberdade, o valor crítico de χ^2 é 5,99. Como o valor calculado é menor do que 5,99, a hipótese da nulidade não deve ser rejeitada.

Se você usar o programa *Statistica* para fazer os cálculos, obterá o *p*-valor, que é 0,6706 > 0,05.

Median Test, Overall Median = 48,00000 (tempconf.sta)

Independent (grouping) variable: FUMO

Chi-Square = ,8000000, df = 2, p = ,6703

		Group 1	**Group 2**	**Group 3**	**Total**
	observed	6,00000	4,00000	5,00000	
<= Median:	expected	5,00000	5,00000	5,00000	15,00000
	obs.-exp.	1,00000	-1,00000	0,00000	
	observed	4,00000	6,00000	5,00000	
> Median:	expected	5,00000	5,00000	5,00000	15,00000
	obs.-exp.	-1,00000	1,00000	0,00000	
Total:	observed	10,00000	10,00000	10,00000	30,00000

Observação: Recomenda-se só aplicar o teste da mediana se a amostra for maior ou igual a 20 e houver pelo menos cinco observações em cada grupo.

4.2. COMPARAÇÃO DE MAIS DE DOIS GRUPOS DEPENDENTES

4.2.1.Teste de Friedman

O teste de Friedman é também conhecido como análise de variância de Friedman, ANOVA de Friedman, ANOVA não paramétrica com dois critérios.

Indicação: O teste de Friedman é indicado para testar a hipótese de que existe diferença entre mais de dois grupos quando esses grupos são dependentes. A variável deve ser ordinal ou numérica.

Para ficar claro o que são grupos dependentes, veja o Exemplo 4.7.

Exemplo 4.7

Imagine que um pesquisador obteve a glicemia de seus pacientes diabéticos antes, durante e depois do tratamento. A observação feita em determinado paciente *antes* do tratamento está relacionada com a observação feita nesse mesmo paciente *durante* o tratamento e com a observação feita *depois* do tratamento. Cada paciente constitui um "bloco" de três dados *dependentes.* Indicamos as três situações por Grupo I (dados de glicemia antes do tratamento), Grupo II (dados de glicemia durante o tratamento) e Grupo III (dados de glicemia depois do tratamento).

Para proceder ao teste de Friedman:

Primeiro passo: A hipótese da nulidade é a de que não há diferença entre grupos. Estabeleça a hipótese alternativa e o nível de significância.

Segundo passo: Atribua um posto a cada dado *por bloco*. Reveja o exemplo dos pacientes diabéticos cuja glicemia foi obtida antes, durante e depois do tratamento. Nesse exemplo, os postos seriam atribuídos por paciente: 1 para o dado com menor valor, 2 para o seguinte, 3 para o dado com maior valor.

Terceiro passo: Calcule as somas dos postos de cada grupo, isto é, $\sum R_1$, $\sum R_2$, ..., $\sum R_k$. Se as somas dos postos estiverem corretas, então:

$$\sum R_1 + \sum R_2 + ... + \sum R_k = \frac{1}{2} nk(k+1) \tag{4.7}$$

Quarto passo: Calcule:

$$\chi_r^2 = \frac{12}{Nk(k+1)} \left(\sum R_1^2 + \sum R_2^2 + ... + \sum R_k^2 \right) - 3N(k+1) \tag{4.8}$$

em que

N = número de unidades (p. ex., pacientes);

k é o número de grupos.

Quinto passo: Sob a hipótese da nulidade, a estatística χ_r^2 tem, aproximadamente, distribuição de $\chi 2$ com (k – 1) graus de liberdade. Faça o teste, que consiste em comparar o valor calculado de χ_r^2 com o valor crítico dado na tabela de $\chi 2$ (Tabela 2 do Apêndice), no nível de significância estabelecido e com (k – 1) graus de liberdade.

Sexto passo: Para comparação, calcule as médias dos postos dos grupos.

Empates

Fique atento. Quando os empates ocorrem em grupos distintos, a perda de poder é maior. Então *não* calcule a estatística χ_r^2 conforme a fórmula (4.5). É preciso correção.[60] Os cálculos se tornam muito trabalhosos, mas muitos programas para computador fazem essa correção automaticamente.

Exemplo 4.8

Para verificar a aprendizagem de seus alunos, um professor de radiologia fez uma perfuração em uma mandíbula usando broca 10. Depois, radiografou a mandíbula usando técnica digital, o que lhe permitiu obter muitas cópias. Pediu, então, a cada um de seus oito alunos que examinassem a radiografia no início, no decorrer e no final do curso. Os alunos não sabiam que examinavam sempre a mesma radiografia e que a resposta correta era “há lesão” (porque havia uma perfuração feita com broca 10 na mandíbula). Nas três ocasiões, cada aluno deveria examinar a radiografia que recebia e dizer:

a) não há lesão;

b) provavelmente não há lesão;

c) provavelmente há lesão;

d) há lesão.

Para a análise, o professor conferiu notas a cada resposta de cada aluno de acordo com o critério:

a) zero: não há lesão;

b) 1: provavelmente não há lesão;

c) 2: provavelmente há lesão;

d) 3: há lesão.

Notas dadas pelos alunos segundo o grupo (período do curso)

Aluno	Grupo		
	Início	Durante	Final
A	0	1	3
B	0	1	2
C	0	2	3
D	1	0	3
E	1	2	3

F	0	3	2
G	0	1	3
H	3	0	1

As notas que você precisa comparar estão relacionadas entre si, porque são dadas pelo mesmo aluno em três períodos distintos do curso. Vamos, então, aplicar o teste de Friedman.

Primeiro passo: A hipótese da nulidade é a de que não há diferença na avaliação dos alunos no início, no decorrer e no final do curso. Seja α = 0,05.

Segundo passo: Atribua um posto a cada dado *por linha*. Os postos, em cada linha, vão de 1 (o dado de menor valor) a 3 (o dado de maior valor). Valores empatados recebem, como posto, a média dos postos que lhes seria conferido.

Terceiro passo: Para cada grupo, calcule a soma dos postos, isto é, $\sum R_1$, $\sum R_2$ e $\sum R_3$. Depois, verifique os cálculos, aplicando (4.5), lembrando que N é o número de alunos (número de linhas) e k é o número de grupos (períodos em que os alunos foram avaliados). Então, N = 8 e k = 3. Para o exemplo:

$$11 + 15 + 22 = \frac{1}{2}(8 \times 3 \times 4) = 48$$

Notas dadas pelos alunos segundo o grupo (Tabela auxiliar para proceder ao teste de Friedman)

Aluno	**Grupo**					
	Início		**Durante**		**Final**	
	Dado	**Posto**	**Dado**	**Posto**	**Dado**	**Posto**
A	0	1	1	2	3	3
B	0	1	1	2	2	3
C	0	1	2	2	3	3
D	1	2	0	1	3	3
E	1	1	2	2	3	3
F	0	1	3	3	2	2
G	0	1	1	2	3	3
H	3	3	0	1	1	2
		$\sum R_1 = 11$		$\sum R_2 = 15$		$\sum R_3 = 22$

Quarto passo: Calcule χ_r^2 conforme a fórmula (4.6), transcrita aqui:

$$\chi^2 = \frac{12}{Nk(k+1)}\left[\left(\sum R_1\right)^2 + \left(\sum R_2\right)^2 \dots \left(\sum R_k\right)^2\right] - 3N(k+1)$$

Neste exemplo, N = 8 alunos (linhas) e k = 3 grupos (colunas).

$$\chi^2 = \frac{12}{8 \times 3 \times (3+1)}[(11)^2 + (15)^2 + (22)^2] - 3 \times 8 \times (3+1)$$

$$\chi^2 = \frac{1}{8}[121 + 225 + 484\ ] - 96 = 7{,}75$$

Quinto passo: Como o valor calculado de χ_r^2 de Friedman é maior do que χ^2 = 5,99, valor crítico com dois graus de liberdade

e no nível de 5% de significância dado na Tabela 2 do Apêndice, o professor pode afirmar que o grau de conhecimento de seus alunos mudou.

Sexto passo: As médias dos postos de cada grupo dão ideia do padrão de diferenças:

- início do curso: média 1,375;
- durante o curso: média 1,750;
- final do curso: média 2,750.

Se você usar o programa *Statistica* para fazer os cálculos, obterá:

Friedman ANOVA and Kendall Coeff. of Concordance (notas.sta)

ANOVA Chi Sqr. (N = 8, df = 2) = 7,750000 p < ,02076

Coeff. of Concordance = ,48438 Aver. rank r = ,41071

	Average	Sum of		
	Rank	**Ranks**	**Mean**	**Std.Dev.**
Antes	1,375000	11,00000	,625000	1,060660
Durante	1,875000	15,00000	1,250000	1,035098
Depois	2,750000	22,00000	2,500000	,755929

Amostras pequenas

Quando as amostras são pequenas, a distribuição da estatística χ_r^2 não se aproxima, satisfatoriamente, da distribuição de $\chi 2$. Se $k = 3$ e N varia entre 2 e 9, ou $k = 4$ e N varia entre 2 e 4, use a Tabela 9 do Apêndice. No entanto, muitos programas para computador usam, mesmo nesses casos, a aproximação normal e fornecem o *p*-valor.

4.2.1.1. Teste de Dunn para comparações múltiplas

Se os dados permitem concluir que as populações amostradas ou os tratamentos estudados não são iguais, é lógico buscar saber onde estão as diferenças. É preciso, então, um procedimento para fazer as comparações múltiplas ou, melhor ainda, comparar os postos médios dois a dois. Nesses casos, é muito usado o teste de Dunn.

Para fazer o teste de Dunn:

Primeiro passo: A hipótese da nulidade é a de que os grupos em comparação têm os mesmos postos médios.[61] Estabeleça a hipótese alternativa e o nível de significância.

Segundo passo: Calcule o desvio padrão *s*

$$s = \sqrt{\frac{bk(k+1)}{6}} \tag{4.9}$$

Nessa fórmula, *b* é o número de unidades (p. ex., pacientes) em comparação e *k* é o número de

grupos.

Terceiro passo: Calcule as estatísticas de teste para comparar as somas dos postos duas a duas:

$$q = \frac{R_i - R_j}{s} \tag{4.10}$$

em que R_i e R_j indicam a soma dos postos dos i-ésimo e j-ésimo grupos, $i = 1, 2,...,k$, $j = 1,2,...k$ e $i \; {}^{4}_{4} \; j$. A diferença entre grupos é significante toda vez que o valor calculado de q for maior do que o valor crítico de q (Tabela 6 do Apêndice), no nível estabelecido de significância e com k grupos.[62] Se você estiver usando um programa para computador, procure o p-valor associado a cada diferença de postos médios.

Exemplo 4.9

Reveja o Exemplo 4.7. Os alunos examinaram a radiografia no início, no decorrer e no final do curso e deram respostas. O conhecimento deles sobre o assunto melhorou?

Primeiro passo: A hipótese em teste é a de que os grupos em comparação têm o mesmo posto médio. Seja 5% o nível de significância.

Segundo passo: Calcule o desvio padrão dos dados conforme a fórmula (4.9):

$$\sqrt{\frac{bk(k+1)}{6}} = \sqrt{\frac{8 \times 3 \times (3+1)}{6}} = 4$$

Terceiro passo: Calcule as estatísticas de teste para comparar as somas dos postos, duas a duas, conforme a fórmula (4.10). Organize os resultados em uma tabela-resumo para o teste de comparações múltiplas. Compare o valor calculado com o valor crítico, para três tratamentos e no nível de significância de 5%.

Tabela-resumo para o teste de comparações múltiplas,exemplo 4.6

Comparação	**$\|R_i - R_j\|$**	**Valor calculado de q**	**Valor crítico de q**	**Conclusão**
Início versus durante	$\|11 - 15\| = 4$	1	2,394	Não rejeita H_0
Início versus final	$\|11 - 22\| = 11$	2,75	2,394	Rejeita H_0
Durante versus final	$\|15 - 22\| = 7$	1,75	2,394	Não rejeita H_0

Os alunos apresentaram melhor conhecimento sobre o assunto quando terminaram o curso, em comparação com o que sabiam no início.

RESUMO E OBJETIVO DO CAPÍTULO

Lido este Capítulo, você deverá ser capaz não só de aplicar os testes apresentados, como também de saber por que foram aplicados. Foram apresentadas duas situações que admitem o uso de estatística não paramétrica:

1. *Comparação de grupos independentes*. Foram apresentados o teste de Kruskal-Wallis e o teste da mediana. As comparações múltiplas podem ser feitas pelo teste de Dunn, que tem no teste de Tukey seu equivalente paramétrico.
2. *Comparação de mais de duas amostras dependentes*. Foi apresentado o teste de Friedman.

---

[49] Veja o procedimento para análise de variância em VIEIRA, S. *Análise de variância*. São Paulo: Atlas, 2006.

[50] Id., ibid.

[51] The Kruskal-Wallis Test for 3 or More Independent Samples. Disponível em: http://vassarstats.net/textbook/ch14a.html. Acesso em 3 de novembro de 2017.

[52] Statsoft, Inc. 2300 East 14th Street, Tulsa, OK 74104, USA.

[53] MCDONALD, JH. *Handbook of Biological Statistics*. 2nd ed. Baltimore: Sparky House Publishing, 2009. p. 165-172.

[54] Alguns dizem que o teste de Dunn é o equivalente não paramétrico do teste de Tukey.

[55] Com base em ZAR, JH. *Biostatistical Analysis*. 4th ed. Upper Saddle River, New Jersey: Prentice Hall, 1999. p. 224-225.

[56] O valor de *q* é o valor de z da distribuição normal padronizada que tem, à sua direita, a área $\sim$/k (k-1). É uma correção baseada na desigualdade de Bonferroni. Veja:

DANIEL, WW. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000. p. 241.

[57] Veja HOLLANDER, M.E.; WOLFE, D.A. *Nonparametric Statistics Methods*. Nova York: Wiley, 1973.

[58] Com base em ZAR, JH. *Biostatistical Analysis*. 4th ed. Upper Saddle River, New Jersey: Prentice Hall, 1999. p. 224-225.

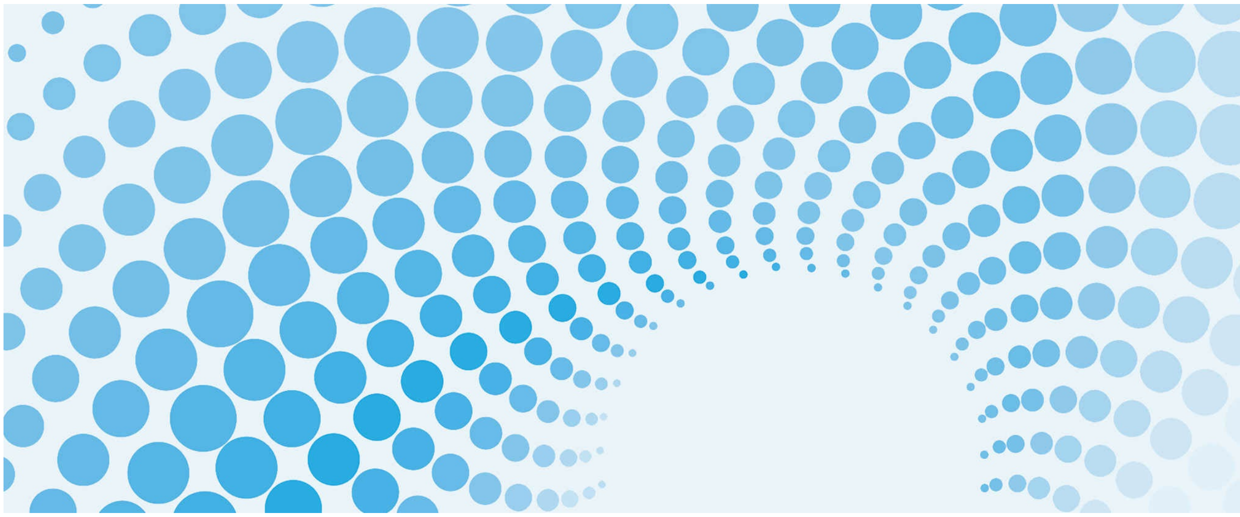
[59] Veja o teste de $\chi 2$ para tabelas 2 × *k* no Capítulo 6 deste livro.

[60] Veja, por exemplo, CONOVER, W.J. *Practical Nonparametric Statistics.* 2nd ed. New York: Wiley, 1980.

[61] Com base em ZAR, J.H. Biostatistical Analysis. 4th ed. Upper Saddle River, New Jersey: Prentice Hall, 1999. p. 224-225.

[62] O valor de *q* é o valor de *z* da distribuição normal padronizada que tem, à sua direita, a área $\sim$/k (k-1). É uma correção baseada na desigualdade de Bonferroni. Veja:

DANIEL, WW. *Applied Nonparametric Statistics*. Pacific Grove: Duxbury, 1990. p. 241.



4.3.1. Homens mais altos têm fibras nervosas mais longas. É possível que o comprimento dos nervos sensoriais esteja diretamente relacionado com a vulnerabilidade. Para saber se as estaturas de homens com pé diabético são maiores, foi medida a estatura de homens de três grupos: não diabéticos (grupo controle), diabéticos e diabéticos com problemas no pé ou grupo de pé diabético. Como a amostra é pequena, compare as estaturas dos três grupos usando o teste de Kruskal-Wallis.

Tabela 4.1 – Estatura, em metros, de não diabéticos (controle),diabéticos e grupo com pé diabético

	Grupo	
Controle	**Diabético**	**Pé diabético**
1,75	1,74	1,77
1,70	1,76	1,89
1,69	1,68	1,84
1,82	1,80	1,81
1,78	1,79	1,73

Fonte: BRESÄTER, L.E., WELIN, L., ROMANUS, B. Foot pathology and risk factors for diabetic foot disease in elderly men. *Diabetes Res Clin Pract 32*(1-2): 103-109, 1996.

4.3.2. São dados[63] os escores de um coeficiente de desempenho verbal para três grupos étnicos. Faça o teste de Kruskal-Wallis.

Tabela 4.2 – Escores do coeficiente de desempenho verbal segundo o grupo étnico

	Etnia	
Negro	**Latino**	**Branco**
1	4	2
3	5	3
6	12	5
7	13	8
9	15	10
10	17	12
11	22	14
16	27	18
20	29	23
21	30	25

Fonte: GIBBONS, J.D. *Nonparametric Statistics*. Newbury Park: Sage Publications, 1993.

4.3.3. Na Tabela 4.3 você encontra o pH de 12 amostras de cada uma de quatro marcas de refrigerante disponíveis no comércio. Cada amostra foi tomada em 1 litro da bebida. Faça o teste da mediana.

Tabela 4.3 – O pH do refrigerante segundo a marca

	Marca		
A	**B**	**C**	**D**
7,1	6,9	7,8	6,4
7,2	7,0	7,9	6,6

7,4	7,1	8,1	6,7
7,6	7,2	8,3	7,1
7,6	7,3	8,3	7,6
7,7	7,3	8,4	7,8
7,7	7,4	8,4	8,2
7,9	7,6	8,4	8,4
8,1	7,8	8,6	8,6
8,4	8,1	8,9	8,7
8,5	8,3	9,2	8,8
8,8	8,5	9,4	8,9

4.3.4. Em uma pesquisa,[64] 10 críticos de arte classificaram quatro pinturas segundo uma escala que variava entre 0 (definitivamente, não gosto) e 5 (definitivamente, gosto muito). A Tabela 4.4 apresenta os valores atribuídos segundo o crítico de arte para cada pintura. Faça o teste de Friedman.

Tabela 4.4 – Valores atribuídos segundo o crítico de arte para cada pintura

Crítico de arte	**Pintura**			
	A	**B**	**C**	**D**
1	0	5	1	4
2	3	4	2	5
3	1	4	3	4
4	4	2	2	3
5	2	2	4	3
6	0	3	5	5
7	3	1	3	4
8	5	3	1	5
9	1	5	2	4
10	2	4	0	3

4.3.5. No exemplo 4.2 é relatado um ensaio clínico randomizado placebo-controlado duplo-cego conduzido com pacientes com osteoartrite para testar o efeito do etoricoxib sobre a dor e comparar o efeito desse medicamento com o do naproxeno. Faça o teste da mediana.

4.3.6. Os postos apresentados[65] em seguida resultam em um teste de Kruskal-Wallis significante. Aplique o teste de Dunn para determinar entre quais grupos populacionais existe diferença estatística.

Grupo A: 8; 4; 3; 5; 1

Grupo B: 10; 6; 9; 11; 2

Grupo C: 14;13;7;12;15

4.3.7. Na Tabela 4.5 são apresentados os níveis de cortisol de três grupos de pacientes cujos partos ocorreram entre 38 e 42 semanas de gestação. O Grupo I foi estudado antes do início da cesárea eletiva, o Grupo II foi estudado durante o trabalho induzido de parto por cesárea e o Grupo III foi estudado durante o trabalho de parto espontâneo. Esses dados trazem

evidência suficiente que permita dizer que existe diferença entre grupos no nível de 1% de significância?

Tabela 4.5 – Nível de cortisol de três grupos de pacientes em trabalho de parto

	Grupo	
I	**II**	**III**
262	465	343
307	501	772
211	455	207
323	355	1.048
454	468	838
339	362	687
304		
154		
287		
356		

Fonte: CAWSON, M.J.; ANDERSON, A.B.M.; TURNBULL, A.C.; LAMPE, L. Cortisol, cortisone, and 11-deoxycortisol levels in human umbilical and maternal plasma in relation to the onset of labour. *J Obstet Gynaecol Br Commonw* 81(10): 737-745, 1974.

4.3.8. A Tabela 4.6 apresenta os escores de desempenho verbal de crianças quando ingressam na escola e vêm de diferentes comunidades. Faça o teste da mediana.

Tabela 4.6 – Escores de desempenho verbal de crianças quando ingressam na escola segundo a comunidade da qual provieram

	Comunidade		
Muito isolada	**Isolada**	**Rural**	**Favela**
21	33	36	41
22	27	35	43
22	29	33	44
20	29	35	37
25	26	39	36
28	29	37	36
22	26	30	42
23	35	33	32
20	33	35	43
44	34	35	25
27	36	27	25
30	40	42	38
30	45	26	34
22	34	36	42
21	26	30	40
25	34	39	45
21	33	30	37

23	33	37	28
26	32	39	24
23	42	33	40
23	28	40	42
28	34	36	41
37	44	41	45

Fonte: DANIEL, W.W. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000. p. 252-253.

4.3.9. Para comparar três métodos de determinação de amilase sérica de pacientes com pancreatite, foram obtidos os valores apresentados na Tabela 4.7. Faça o teste de Friedman e as comparações múltiplas.

Tabela 4.7 – Valores de amilase sérica (unidades de enzima por 100 mL de soro)de pacientes com pancreatite

Paciente	**Método de determinação**		
	A	**B**	**C**
1	4.000	3.210	6.210
2	1.600	1.040	2.410
3	1.600	647	2.210
4	1.200	570	2.060
5	840	445	1.400
6	352	156	249
7	224	155	224
8	200	99	208
9	184	70	227

Fonte: HALL, F.F. et al. An improved amylase assay using a new starch derivative. *Am J Clin Pathol 53*(5): 627634, 1979. Apud DANIEL, W.W. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000. p. 265.

4.3.10. A Tabela 4.8 apresenta os valores de bilirrubina em 10 bebês normais, medidos em idades diferentes. Esses dados trazem evidência suficiente que permita dizer que o nível de bilirrubina decresce entre os 4 e os 10 dias de idade? Faça o teste de Friedman.

Tabela 4.8 – Valores de bilirrubina em 10 bebês normais segundo a idade

Número	**Grupo de idade (em dias)**		
	4	**7**	**10**
1	10,8	5,0	2,6
2	12,5	11,0	6,0
3	13,7	15,6	10,6
4	11,5	3,5	1,6
5	10,2	3,0	2,0
6	8,0	7,0	3,0
7	10,8	7,0	5,6
8	14,9	9,4	7,6
9	16,2	10,2	7,4

10	10,8	4,6	3,5

Fonte: STERN, L., KHANNA, N.N., LEVEY, G., YAFFE, S.J. Effect of phenobarbitol on hyperbilirubinemia and glucuronide formation in newborns. *Am J Dis Child* 120 (1): 26-31, 1970. Apud DANIEL, W.W. *Applied Nonparametric Statistics*. 2nd ed. Pacific Grove: Duxbury, 2000.

---

[63] GIBBONS, J.D. *Nonparametric Statistics*. Newbury Park: Sage Publications, 1993.

[64] Disponível em https://secure.brightstat.com/index.php?p=c&d=1&c=2o. Acess em 6 de novembro de 2017.

[65] ZAR, J.H. *Biostatistical Analysis*. 4th ed. Upper Saddle River, New Jersey: Prentice Hall, 1999. p. 230.

5

Tabelas 2 × 2

Uma *tabela de contingência* (também conhecida como tabela cruzada ou tabela de dupla entrada) é um arranjo de dados classificados de acordo com duas variáveis qualitativas ou categorizadas. Uma das variáveis é apresentada nas linhas e a outra, nas colunas. Cada célula apresenta a frequência de casos para um par específico de categorias de cada uma das variáveis.

Exemplo 5.1

Imagine que um pesquisador quer saber se a opção de jogar lixo nas ruas ou levar o lixo às lixeiras disponíveis em locais públicos está associada ao sexo. Esse pesquisador observa 100 pessoas e registra o sexo e a opção para a forma como essa pessoa descarta o lixo. Os dados obtidos são apresentados em uma tabela, denominada tabela de contingência.

Distribuição das pessoas segundo o sexo e a opção para descarte de lixo

Sexo	**Descarte de lixo**	
	Lixeira	**Rua**
Mulheres	18	7
Homens	42	33

Neste capítulo, vamos estudar as tabelas 2 × 2, que são uma das maneiras mais comuns de apresentar evidência estatística. No caso dessas tabelas, é usual aplicar o teste de qui-quadrado para testar a *hipótese de associação* das variáveis (como no Exemplo 5.1, saber se a opção para a forma de descartar o lixo está associada ao sexo) ou a *hipótese de igualdade de proporções*, como veremos adiante.

As hipóteses são redigidas de maneira diferente porque as frequências exibidas nas tabelas 2 × 2 são obtidas por diferentes procedimentos de pesquisa: estudos transversais, estudos retrospectivos, estudos prospectivos e ensaios clínicos. Vamos entender esses procedimentos por meio de exemplos.

Estudos transversais são feitos para testar a hipótese de que duas variáveis estão associadas. Pesquisadores obtêm uma amostra de tamanho n de uma grande população e classificam cada indivíduo amostrado segundo as duas variáveis de interesse, simultaneamente. Quando essas variáveis são apenas duas (p. ex., as características A e B) e binárias (Sim e Não), os dados são obtidos como mostra esquematicamente a Figura 5.1.

Na apresentação dos dados obtidos em estudos transversais, é recomendável fornecer não apenas as frequências, mas também as proporções – ou as porcentagens – em cada célula. As proporções são obtidas dividindo-se as frequências observadas pelo tamanho da amostra, indicado por n. Para obter a porcentagem, é só multiplicar a proporção por 100.

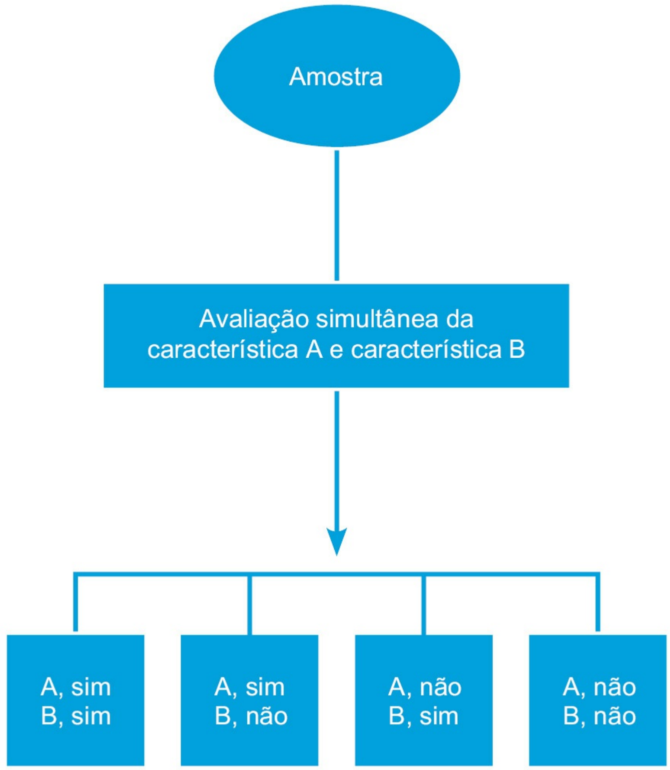


Figura 5.1 – Estudo transversal

Exemplo 5.2

Perguntou-se, a cada uma de 200 crianças de pré-escola, se elas gostavam mais de gato ou de cachorro e, com a resposta, anotou-se o sexo da criança. As respostas foram, portanto, classificadas de acordo com duas variáveis categorizadas: sexo (masculino ou feminino) e animal preferido (gato ou cachorro). Os dados obtidos foram:

- 51 meninas preferiam gato.
- 42 meninas preferiam cachorro.
- 50 meninos preferiam gato.
- 57 meninos preferiam cachorro.

Esses dados são arranjados em uma *tabela de contingência.*

Sexo	**Animal preferido**	
	Gato	**Cachorro**
Masculino	50 (0,250)	57 (0,285)
Feminino	51 (0,255)	42 (0,210)

Nos *estudos retrospectivos*, os participantes provêm de duas populações diferentes: os que têm determinada característica (casos) e os que não têm essa característica (controles). Uma amostra de pessoas de cada população é examinada para determinar a proporção dos que foram

expostos a determinado fator que se presume de risco.

Por exemplo, pesquisadores amostram n_1 pessoas com uma doença cardiovascular e n_2 pessoas sem essa doença. Os dois grupos devem ser em tudo comparáveis exceto pelo fato de, em um deles, as pessoas terem a doença e no outro, não. Depois, os pesquisadores estudam a história pregressa das pessoas para obter a proporção, tanto no grupo de doentes como no grupo de não doentes, de pessoas que foram expostas a um fator que eles presumem de risco, como o sedentarismo, por exemplo. Veja a Figura 5.2.

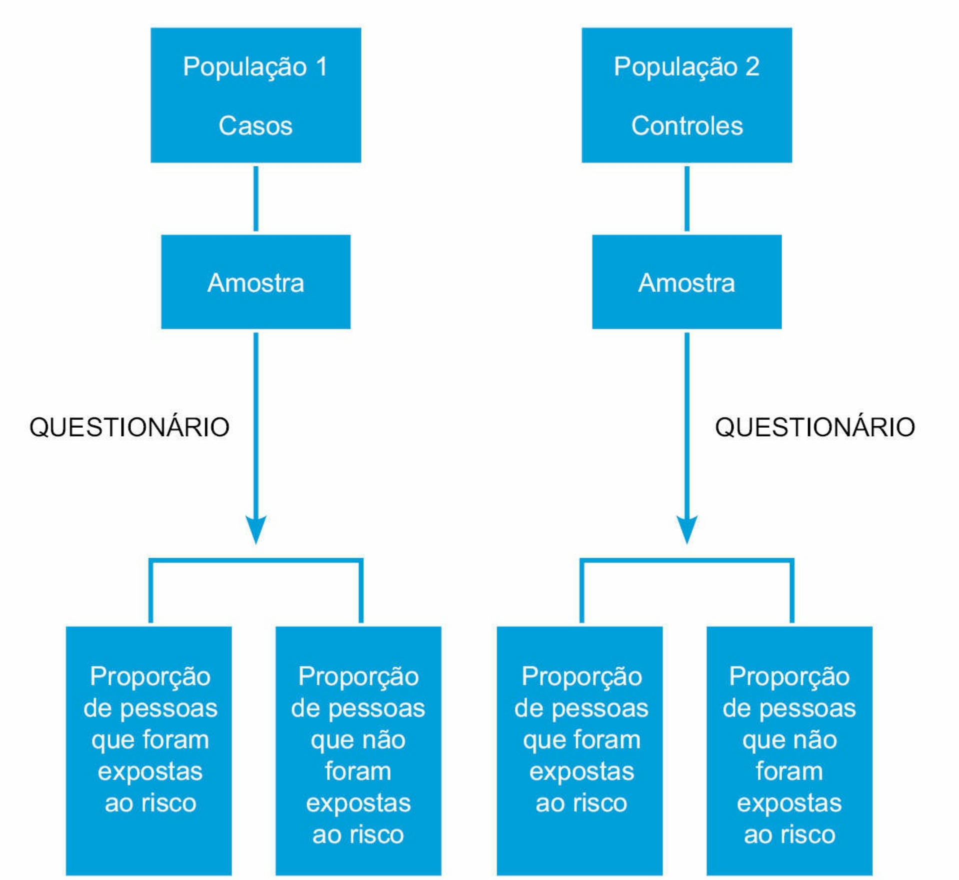


Figura 5.2 – Estudo retrospectivo.

Nos estudos retrospectivos, é recomendável fornecer as proporções (ou porcentagens) que estão sendo comparadas já na apresentação dos dados. Veja o exemplo 5.3, que é clássico, porque foi a primeira vez que se apresentou esse procedimento de pesquisa.

Exemplo 5.3

Foi feito um estudo retrospectivo para saber se câncer no pulmão ocorre mais em fumantes. Foram entrevistados dois grupos de pessoas que estavam sendo atendidas em hospitais de Londres e cidades vizinhas para conhecer seus hábitos de fumar: um grupo de 649 pessoas tinha câncer no pulmão (os casos) e outro grupo, de 649 pessoas comparáveis, não tinha câncer de pulmão (os controles). Dos casos, 27 eram fumantes; dos controles, 2 eram fumantes.

Participantes da pesquisa segundo ter ou não câncer de pulmão e ser ou não fumante

Câncer no pulmão	Hábito de fumar		Total	Porcentagem de fumantes
	Sim	Não		
Presente	27	622	649	4,16
Ausente	2	647	649	0,308
Total	29	1.269	1.298	

Fonte: DOLL, R.; HILL, A.B. Smoking and carcinoma of the lung. *Br Med J* 2: 739-48, 1950.

Os estudos retrospectivos são pouco confiáveis porque neles se pede que as pessoas relatem fatos do passado, que podem ser sobrevalorizados se as pessoas suspeitarem de que tais fatos explicam as condições em que elas estão no presente. Para compensar essa desvantagem, recomenda-se que esses estudos sejam feitos com grandes amostras. No entanto, isso demanda tempo e dinheiro. Então os estudos retrospectivos são, muitas vezes, feitos com poucos dados, o que os torna rápidos e baratos – mas, por conta disso, pouco confiáveis.

Nos *estudos prospectivos*, também chamados coortes, os participantes provêm de duas populações diferentes: os que estão expostos a determinado fator que se presume de risco para determinado desfecho e os que não estão expostos a esse fator. Pesquisadores acompanham uma amostra de n_1 pessoas expostas a esse fator e uma amostra de n_2 pessoas não expostas, durante tempo relativamente longo. Terminado o período de observação, calcula-se a proporção de pessoas com o desfecho procurado em cada uma das amostras. Veja a Figura 5.3.

Como exemplo, imagine que pesquisadores amostram n_1 pessoas sedentárias e n_2 pessoas não sedentárias para saber se o sedentarismo aumenta o risco de doença cardiovascular. Os dois grupos devem ser em tudo comparáveis exceto pelo fato de, em um deles, as pessoas serem sedentárias e no outro, não. Depois de um período relativamente longo, os pesquisadores determinam a proporção de pessoas com a doença nas duas amostras.

Nos estudos prospectivos é recomendável apresentar, nas tabelas 2 × 2 que contêm os dados, as proporções que estão sendo comparadas. Veja o Exemplo 5.4.

Exemplo 5.4

Pesquisadores acompanharam dois grupos de mulheres em período de gestação até o parto: o primeiro grupo era constituído por diabéticas e o segundo, por não diabéticas, pois se suspeita que o óbito neonatal ocorra mais quando a mãe é diabética. Os pesquisadores contaram o número de casos de óbito neonatal tanto no grupo de diabéticas como no grupo de não diabéticas. Os dados obtidos e as proporções em comparação estão apresentados em uma tabela 2 × 2.

Óbito neonatal segundo o fato de a mãe ter ou não diabetes melito

Diabetes melito	Óbito neonatal		Total	Porcentagem de óbitos neonatais
	Sim	Não		
Portadoras	3	21	24	12,50%

Não portadoras	21	830	851	2,47%
Total	24	851	875	

Fonte: CUNHA, A.A, PORTELA, M.C., AMED, A.M., CAMANO, L. *Diabetes mellitus* tipo 1 (insulinodependente) e gravidez: conduta obstétrica e resultado perinatal. *Go Atual 5*(6): 24-26, 2001.

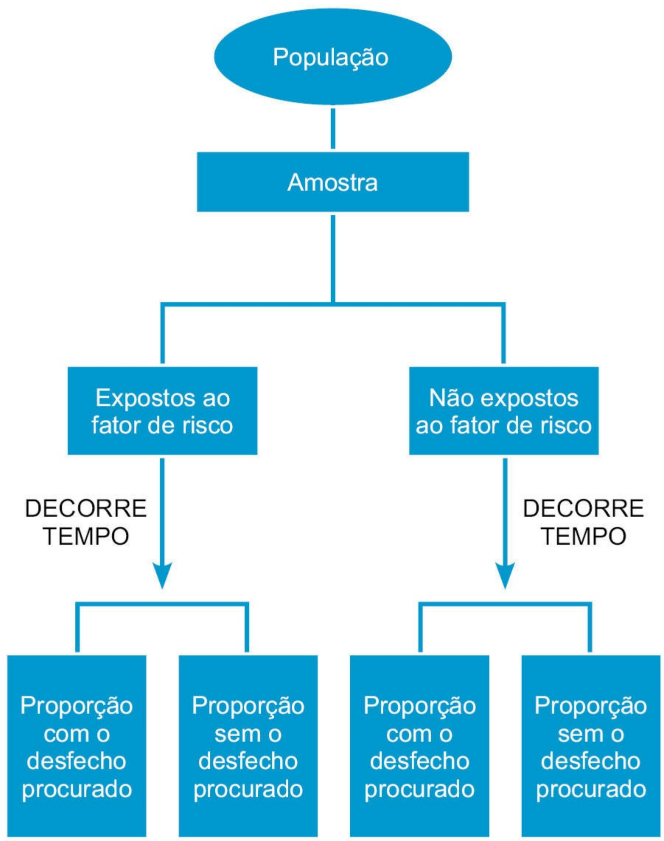


Figura 5.3 – Estudo prospectivo.

Para fazer um ensaio clínico randomizado e controlado, o pesquisador divide uma amostra de n pacientes *ao acaso* em dois grupos: um grupo de n_1 pacientes recebe o tratamento em teste (grupo tratado) e o outro, de n_2 pacientes, constitui o controle (controle positivo se os pacientes receberem um tratamento convencional e controle negativo se os pacientes receberem um placebo). Se o desfecho do tratamento só puder ser "sim" ou "não", os dados do ensaio são apresentados em uma tabela 2 × 2. Veja a Figura 5.4.

Como os ensaios clínicos são feitos para *comparar grupos*, é recomendável fornecer as proporções com respostas positivas dos dois grupos quando se apresentam os dados. Veja o Exemplo 5.5.

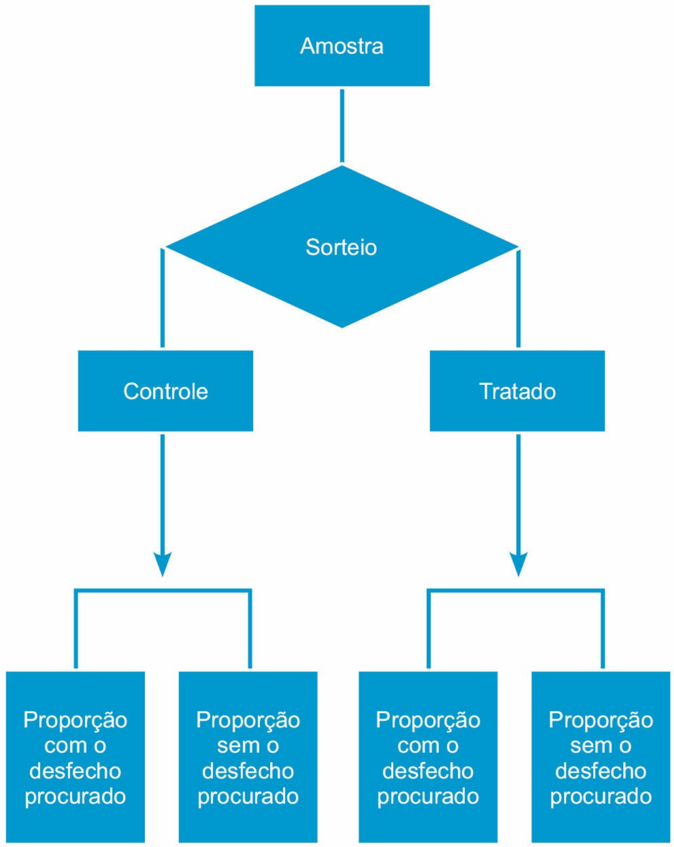


Figura 5.4. Estudo clínico.

Exemplo 5.5

Pacientes em UTI com ventilação mecânica correm risco de desenvolver pneumonia aspirativa, causada pelo refluxo do alimento e pela consequente aspiração. Em um ensaio clínico para verificar se a nutrição enteral estéril diminui a colonização microbiana do estômago do paciente em ventilação mecânica, o grupo controle recebeu dieta enteral regular e o grupo tratado recebeu a dieta em teste. Dos 20 tratados, 3 tiveram pneumonia aspirativa; dos 20 controles, 2 tiveram pneumonia aspirativa.

Incidência de pneumonia aspirativa segundo o grupo

Grupo	**Pneumonia aspirativa**		**Total**	**Proporção com pneumonia**
	Sim	**Não**		
Tratado	3	17	20	0,150
Controle	2	18	20	0,100
Total	5	35	40	

Fonte: DAVID, C.M.; GOLDWASSER, R.; GONTIJO FILHO, P.P. A nutrição enteral estéril diminui a colonização microbiana do estômago do paciente em ventilação mecânica. *Rev Bras Terap Intens 10*(1): 12-28, 1998.

5.1. TESTE DE χ2 DE PEARSON

Indicação: O teste de χ2 de Pearson serve para testar a hipótese de que *duas variáveis categorizadas são independentes.*[66] O símbolo χ é uma letra grega de nome qui;[67] lê-se teste de qui-quadrado porque χ está elevado ao quadrado (à segunda potência).

Para fazer o teste de χ2 de Pearson:

Primeiro passo: Escreva a hipótese da nulidade, de que as variáveis são independentes. Estabeleça o nível de significância.

Segundo passo: Apresente os dados em uma tabela 2 × 2. Calcule os *totais marginais* e o total geral, como mostrado na Tabela 5.1. Os totais estão em negrito.

Tabela 5.1 – Valores literais em uma tabela 2 × 2

Variável A	**Variável B**		**Total**
	Não	**Sim**	
Não	a	b	a + b
Sim	c	d	c + d
Total	a + c	b + d	n

Terceiro passo: Calcule a estatística de teste, que está associada a 1 grau de liberdade.

$$\chi^2 = \frac{(ad - bc)^2\, n}{(a+b)(c+d)(a+c)(b+d)} \tag{5.1}$$

Quarto passo: Compare o valor calculado de χ2 que está associado a 1 grau de liberdade com o valor crítico, no nível de significância estabelecido (Tabela 2 do Apêndice). Rejeite a hipótese da nulidade toda vez que o valor calculado de χ2 for igual ou maior do que o valor crítico. Se você estiver usando um programa de computador para fazer os cálculos, vai obter o *p*-valor.

Exemplo 5.6

Para estudar a associação entre correr mais de 25 km/semana e ter dor nos joelhos, imagine que pesquisadores fizeram um estudo transversal com 1.455 universitários. Dos 1.000 que não tinham dor nos joelhos, 215 não corriam mais de 25 km/semana. Dos 455 que tinham dor no joelho, 75 não corriam mais de 25 km/semana. Para fazer o teste de χ2 de Pearson:

Primeiro passo: A hipótese da nulidade é a de que a dor nos joelhos *não depende* de correr mais de 25km/semana. Estabeleça o nível de significância de 0,05.

Segundo passo: Apresente os dados em uma tabela 2 × 2 e calcule os *totais marginais* e o *total geral*.

Participantes da pesquisa segundo o fato de ter dor nos joelhos e correr mais de 25 km/semana

Corre mais de	**Tem dor nos joelhos**	**Total**

25 km /semana	Não	Sim	
Não	215	75	290
Sim	785	380	1.165
Total	1.000	455	1.455

Fonte: http://www.theanalysisfactor.com/difference-between-chi-square-test-and-mcnemar-test/

Terceiro passo: Calcule a estatística de teste usando a fórmula (5.1):

$$\chi^2 = \frac{(215 \times 380 - 75 \times 785)^2 \times 1455}{(215 + 75) \times (785 + 380) \times (215 + 785) \times (75 + 380)} =$$

$$\frac{(22825)^2 \times 1.455}{290 \times 1.165 \times 1.000 \times 455} = \frac{758026809375}{153721750000} = 4{,}931$$

Quarto passo: Compare o valor calculado de χ2 (4,931), que está associado a 1 grau de liberdade, com o valor crítico de χ2 no nível de significância de 0,05 (Tabela 2 do Apêndice). O valor crítico é 3,84. Como o valor calculado de χ2 é maior do que o valor crítico, a conclusão é a de que, no nível de 5% de significância, dores no joelho estão associadas ao fato de correr mais de 25 km/semana.

Se você estiver usando um programa de computador para fazer esses cálculos, vai encontrar *p*-valor = 0,026377.

Exigências para a aplicação do teste

1. *Independência*: as características estudadas devem ser independentes, como, por exemplo, um grupo é constituído por homens, outro por mulheres e se pergunta a eles se fumam ou não fumam.
2. *Tamanho da amostra*:[68] a amostra deve ter tamanho maior do que 20 (n > 20) e as frequências marginais não podem ser pequenas.

5.1.1. Correção de continuidade

Muitos estatísticos recomendam, nos casos de uma tabela 2 × 2, calcular o valor de χ2 de Pearson *com correção de continuidade.* A estatística, conhecida como χ2 *com correção de Yates*, é:

$$\chi^2 = \frac{\left(|ad - bc| - \frac{n}{2}\right)^2 n}{(a+b)(c+d)(a+c)(b+d)} \tag{5.2}$$

e está associada a (2 –1) (2 – 1) = 1 grau de liberdade.

A correção de continuidade produz um teste mais *conservador*, isto é, um teste com *menor* probabilidade de rejeitar a hipótese da nulidade.[69] Se a amostra for pequena, o efeito da correção de continuidade será ainda maior.

Os programas para computador fornecem o valor de χ2 tanto para o teste com correção de continuidade como sem correção, bem como os respectivos *p*-valores. E preste muita atenção porque isto acontece: você aplica o teste de χ2 para testar a independência de duas variáveis a determinado conjunto de dados e, sem a correção de continuidade, o resultado é significante, mas, com a correção, é não significante. Uma opção razoável – desde que a literatura confirme seus resultados – é aplicar o teste de χ2 sem a correção de continuidade.

Exemplo 5.7

Para o exemplo 5.6, você também pode calcular o valor de χ2 com correção de continuidade. Aplicando a fórmula (5.2), vem:

$$\chi^2 = \frac{\left(|215 \times 380 - 75 \times 785\ | - \frac{1.455}{2}\right)^2 \times 1.455}{(215 + 75\,) \times (785 + 380\,) \times (215 + 785\,) \times (75 + 380)} =$$

$$\frac{(22097{,}5)^2 \times 1455}{290 \times 1.165 \times 1.000 \times 455} = \frac{710475781593{,}75}{153721750000} = 4{,}622$$

Como o valor calculado de χ2 (4,622) é maior do que o valor crítico com 1 grau de liberdade e no nível de significância de 5% (3,84), a conclusão é a de que, nesse nível de significância, dores no joelho estão associadas ao fato de correr mais de 25 km/semana.

Se você estiver usando um programa para computador, vai obter o *p*-valor = 0,0316 (maior do que o obtido anteriormente, porque o teste com correção de continuidade se torna mais conservador).

5.2. TESTE DE PROPORÇÕES POPULACIONAIS

Os estudos prospectivos e os ensaios clínicos são feitos para comparar proporções populacionais ou probabilidades.[70]

Exemplo 5.8

Para comparar tratamento não operatório com cirurgia conservadora nos casos de trauma esplênico, 136 pacientes foram randomizados para essas duas intervenções. O desfecho era haver ou não haver complicações. Dos 32 não operados, 3 tiveram complicações e, dos 104 submetidos à cirurgia, 25 tiveram complicações.[71] Foram, então, comparadas as proporções de pacientes com complicações nos dois grupos.

Para proceder ao teste:

Primeiro passo: Estabeleça a hipótese da nulidade de que a probabilidade de unidades com determinada característica é a mesma nos dois grupos. A hipótese alternativa é de que existe diferença entre grupos. Estabeleça o nível de significância.

Segundo passo: Calcule as *frequências marginais* e o *total geral*, como mostra a Tabela 5.2.

Tabela 5.2 – Valores literais em uma tabela 2 × 2

Grupo	**Característica A**		
	Sim	**Não**	**Total**
1	a	b	a + b
2	c	d	c + d
Total	a + c	b + d	n

Terceiro passo: Calcule a proporção de unidades com a característica A no Grupo 1:

$$p_1 = \frac{a}{a+b}$$

Quarto passo: Calcule a proporção de unidades com a característica A no Grupo 2:

$$p_2 = \frac{c}{c+d}$$

Quinto passo: Calcule a proporção de unidades *com* a característica A nos dois grupos:

$$p = \frac{a+c}{n}$$

Sexto passo: Calcule a proporção de unidades *sem* a característica A nos dois grupos:

$$q = \frac{b+d}{n}$$

Verifique que:

$$p + q = 1$$

Sétimo passo: Calcule a estatística:

$$z = \frac{p_2 - p_1}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{5.4}$$

Oitavo passo: Sob a hipótese da nulidade, a estatística z tem distribuição aproximadamente normal padronizada,[72] desde que $np > 5$ e $nq > 5$. Compare o valor absoluto de z calculado com o valor crítico de z (Tabela 1 do Apêndice) no nível estabelecido de significância. Se você estiver usando um programa para computador para fazer esses cálculos, vai obter o p-valor.

Exemplo 5.9

Reveja o Exemplo 5.8. Para fazer o teste:

Primeiro passo: Hipótese da nulidade – a probabilidade de pacientes *que têm complicações* quando submetidos a tratamento não operatório é igual à probabilidade de pacientes *que têm complicações* quando submetidos à cirurgia conservadora. Seja $\alpha = 0{,}05$.

Segundo passo: Apresente os dados em uma tabela 2 × 2. Calcule as frequências marginais e o total geral.

Participantes da pesquisa segundo o tipo de tratamento e a ocorrência ou não de complicações

Tratamento	**Complicações**		**Total**
	Sim	**Não**	
Não operatório	3	29	32
Cirúrgico	25	79	104
Total	28	108	136

Fonte: SCARPELINl, S.; ANDRADE, J.I.; STRACIERE, L.D.S.; GRADE, M.H.C.; MACCHETI, A.H.; PASSOS, ADC.Estudo comparativo entre o tratamento não operatório e a cirurgia conservadora no trauma esplênico. *Rev Col Bras Cir* 26(5): 281-284, 1999.

Terceiro passo: Calcule as proporções, que estão dadas na tabela.

Participantes da pesquisa segundo o tipo de tratamento e a ocorrência ou não de complicações (tabela auxiliar)

Tratamento	**Complicações**		**Total**	**Proporção (sem complicações)**
	Sim	**Não**		
Não operatório	3	29	32	0,093750
Cirúrgico	25	79	104	0,240385
Total	28	108	136	

Proporção	0,205882	0,794118

Quarto passo: Calcule a estatística de teste, dada pela fórmula (5.4):

$$z = \frac{0{,}093750 - 0{,}240382}{\sqrt{0{,}205882 \times 0{,}794118 \times \left(\frac{1}{32} + \frac{1}{104}\right)}} = -1{,}794$$

Quinto passo: Como o valor absoluto de *z* calculado (–1,794) é, em valor absoluto, menor do que o valor crítico no nível de 5% de significância (1,96), não se rejeita a hipótese de que, nos casos de trauma esplênico, a probabilidade de pacientes com complicações é a mesma, quer se faça tratamento não operatório, quer se faça cirurgia conservadora.

Se você estiver usando um programa de computador para fazer esses cálculos, vai obter *p*-valor = 0,072813.

5.2.1. Correção de continuidade

Muitos estatísticos recomendam calcular o valor de *z com correção de continuidade*. A fórmula (3.4) fica, então, como segue:

$$z = \frac{|p_2 - p_1| - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{5.5}$$

em que |*p*2 – *p*1| indica o valor absoluto da diferença.

O valor de *z* com a correção de continuidade é sempre menor do que o valor de *z* sem a correção. Logo, a correção de continuidade produz um teste mais *conservador*, isto é, um teste que tem menor probabilidade de rejeitar a hipótese da nulidade. Se a amostra for pequena, o efeito da correção de continuidade será ainda maior.

IMPORTANTE:

No caso de uma tabela 2 × 2, a fórmula (5.2) é matematicamente igual ao quadrado da fórmula (5.4).

As distribuições estatísticas, neste caso, são iguais.[73]

Entretanto, as hipóteses em teste são diferentes.[74]

5.3. TESTE EXATO DE FISHER

Indicação: Você usa o teste exato de Fisher quando quer testar a hipótese de que a probabilidade de ocorrer determinada categoria de uma variável é a mesma em dois grupos. Além disso, a amostra é pequena.

Exemplo 5.10

Em uma pesquisa se perguntou a um grupo de pessoas se gostariam de assistir a determinado jogo de futebol no estádio. O objetivo era testar a hipótese de nulidade de que é menos provável um homem dizer "Não", do que uma mulher. Veja os dados na Tabela 5.3.

Tabela 5.3 – Participantes de pesquisa segundo o sexo e a resposta

Sexo	**Resposta**		**Total**
	Sim	**Não**	
Homens	9	1	**10**
Mulheres	5	4	**9**
Total	**14**	**5**	**19**

Os cálculos necessários para proceder ao teste exato de Fisher são muito complicados. Isso não seria obstáculo para os pesquisadores de hoje, que possuem computadores, mas é a razão de esse teste ser menos conhecido do que o teste de χ2 de Pearson.

De qualquer forma, o teste exato de Fisher apresenta outra dificuldade: *não* usa uma função matemática que permita calcular um valor, que seria a estatística de teste. É preciso calcular a *probabilidade* de obter os dados observados e, além disso, as *probabilidades* de obter, sob a hipótese da nulidade, *valores mais extremos do que os observados*. Essas probabilidades são obtidas por meio da fórmula:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \tag{5.3}$$

Calcular essas probabilidades é trabalho demorado, por conta dos fatoriais[75] (indicados pelo símbolo "!"). No entanto, para tornar a questão mais concreta, vamos estudar o procedimento do teste por meio de exemplos. Reveja o Exemplo 5.10. Procure na tabela a célula com menor frequência: é 1. Vamos calcular a probabilidade de ocorrer esse valor usando a fórmula (5.3).

$$P(1) = \frac{(9+1)!\,(5+4)!\,(9+5)!\,(1+4)!}{19!\,9!\,1!\,5!\,4!} = 0{,}1084$$

O que seria mais “extremo” do que, entre 10 homens, apenas um dizer “Não” querer ir ao estádio para assistir a uma partida de futebol? Seria, entre 10 homens, *nenhum* dizer “Não”. Como a amostra já foi tomada, as frequências marginais precisam ser mantidas. No entanto, podemos calcular a probabilidade de *nenhum homem*, em 10, dizer “Não”. Veja como ficaria a distribuição dos dados:

Participantes de pesquisa segundo sexo e opinião (Tabela auxiliar)

Sexo	Opinião		Total
	Sim	**Não**	
Homens	10	0	**10**
Mulheres	4	5	**9**
Total	**14**	**5**	**19**

A probabilidade de ocorrer 0 (zero), valor mais extremo, é:

$$P(0) = \frac{(10+0)!\,(4+5)!\,(10+4)!\,(0+5)!}{19!\,10!\,0!\,4!\,5!} = 0{,}0108$$

Vamos sistematizar o procedimento para o teste exato de Fisher.

Primeiro passo: Estabeleça a hipótese da nulidade, de que a probabilidade de respostas negativas não difere entre os sexos. A hipótese alternativa é a de que a probabilidade de respostas negativas é menor entre homens. Seja $\alpha = 5\%$. Note: o teste é *unilateral*.

Segundo passo: Procure, na tabela que apresenta os dados, a célula com menor frequência. Calcule a probabilidade de obter os *valores observados* nessa tabela usando a fórmula (5.3).

Terceiro passo: Você tem a célula com menor frequência. Construa, agora, *outras tabelas com as mesmas frequências marginais*, diminuindo sucessivamente o valor numérico da célula que apresenta a menor frequência. No Exemplo 5.10, a menor frequência é 1. Então construa outra tabela com frequência 0 no lugar da frequência 1, mantendo *fixos* os valores dos totais marginais da tabela. Depois calcule a probabilidade de obter esse *valor extremo*.

Entende-se *valores extremos* em termos de probabilidade. Valores extremos são aqueles que *poderiam* ocorrer, mas têm *probabilidade igual ou menor* do que a dos valores observados. No exemplo, mais “extremo” do que, entre 10 homens, 1 “Não” querer ir ao estádio seria, em 10 homens, ninguém dizer “Não”. Calculam-se as probabilidades de ocorrer tanto zero como um homem dizendo “Não”, ou seja, $P(0)$ e $P(1)$.

Quarto passo: Some as probabilidades calculadas. Para o Exemplo 5.10:

$$0{,}1084 + 0{,}0108 = 0{,}1192$$

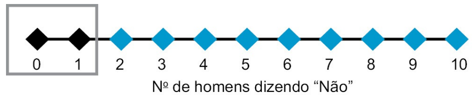


Figura 5.5 Valores correspondentes ao *p*-valor.

Quinto passo: O procedimento do teste consiste em rejeitar a hipótese da nulidade toda vez que a soma das probabilidades calculadas *for igual ou menor do que o nível de significância estabelecido*. A soma das probabilidades calculadas é o *p*-valor. No exemplo, não se rejeita a hipótese de que a probabilidade de respostas negativas seja a mesma para os dois sexos, uma vez que o *p*-valor é 0,1192 > 0,05.

Exemplo 5.11

Imagine que se perguntou, a 12 participantes de pesquisa, se já haviam sido vacinados contra gripe e se tiveram ou não gripe no inverno seguinte. As respostas foram: de 7 vacinados 1 teve gripe e, de 5 não vacinados, 4 tiveram gripe no ano anterior. O pesquisador quer saber se a probabilidade de ter gripe é menor no grupo vacinado. Não pode optar pelo teste de $\chi 2$ porque a amostra é muito pequena. Precisa fazer o teste exato de Fisher.

Participantes da pesquisa classificados segundo as características: ter sido vacinado contra gripe e ter tido gripe no inverno seguinte

Ter sido vacinado	**Ter tido gripe**		**Total**
	Sim	**Não**	
Sim	1	6	**7**
Não	4	1	**5**
Total	**5**	**7**	**12**

Primeiro passo: A hipótese da nulidade é a de que a probabilidade de ter gripe é a mesma nos dois grupos. A hipótese alternativa é a de que a probabilidade de ter gripe é menor no grupo vacinado. O nível de significância é $\alpha = 0,05$.

Segundo passo: Os dados estão em uma tabela 2×2; foram calculados os totais marginais e o total geral. A menor frequência é 1. Vamos trabalhar com a primeira célula, “Sim”-“Sim”. A probabilidade de obter os valores observados usando a fórmula (5.3) é:

$$P(1) = \frac{(1+6)!\,(4+1)!\,(1+4)!\,(6+1)!}{12!\,1!\,6!\,4!\,1!} = 0{,}044192$$

Terceiro passo: Vamos construir outra tabela, com resultado mais extremo do que o obtido e os mesmos totais marginais (em negrito nas duas tabelas). Como? De 7 vacinados, 1 teve gripe. Então seria mais extremo, de sete vacinados, nenhum ter tido gripe.

Participantes da pesquisa classificados segundo as características:ter sido vacinado contra gripe e ter tido gripe no inverno seguinte Tabela auxiliar: valores mais extremos

Ter sido vacinado	**Ter tido gripe**		**Total**
	Sim	**Não**	
Sim	0	7	**7**
Não	5	0	**5**

Total	**5**	**7**	**12**

Calcule a probabilidade de obter os valores mais extremos, apresentados na tabela, usando a mesma fórmula (5.3):

$$P(0) = \frac{(0+7)!\,(5+0)!\,(0+5)!\,(7+0)!}{12!\,0!\,7!\,5!\,0!} = 0{,}00126$$

Quarto passo: Some as probabilidades calculadas.

$$0{,}04419 + 0{,}00126 = 0{,}04545$$

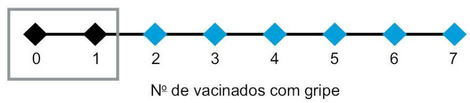


Figura 5.6 Valores correspondentes ao *p*-valor.

Quinto passo: Como o *p*-valor (soma das probabilidades calculadas) é 0,04545 < 0,05, existe evidência de que a probabilidade de ter gripe é menor no grupo vacinado.

O teste exato de Fisher, como feito aqui, é unilateral. Para fazer um teste bilateral, o *p*-valor não pode ser, simplesmente, duplicado, porque a distribuição pode ser assimétrica. É preciso calcular as probabilidades associadas aos valores mais extremos dos dois lados.

Exemplo 5.12

Um pesquisador quer saber se a proporção de pessoas que se queixam de dor é a mesma nos dois sexos. Para 15 pacientes de uma clínica odontológica que estavam sendo submetidos ao mesmo tratamento, foram anotados o sexo e o fato de ter ou não feito queixa de dor durante o tratamento. Como a amostra é muito pequena, deve ser aplicado o teste exato de Fisher.[76]

Pacientes classificados segundo o sexo e a queixa de dor

Sexo	**Queixa de dor**		**Total**
	Sim	**Não**	
Mulher	2	3	**5**
Homem	6	4	**10**
Total	**8**	**7**	**15**

Primeiro passo: A hipótese da nulidade é a de que a proporção de pessoas que se queixam de dor é a mesma nos dois sexos. A hipótese alternativa é a de que a proporção de pessoas que se queixam de dor é diferente nesses dois grupos. Seja $\alpha = 5\%$.

Segundo passo: Calcule a probabilidade de ocorrerem os valores observados usando a fórmula (5.3):

$$P(2,S) = \frac{(2+3)!\,(6+4)!\,(2+6)!\,(3+4)!}{15!\,2!\,3!\,6!\,4!} = 0{,}326$$

Terceiro passo: Calcule a probabilidade de ocorrerem valores mais extremos do que os observados usando a fórmula (5.3). Para

isso, é preciso construir outras tabelas com os mesmos totais marginais, porém com resultados “mais extremos” do que os observados.

A menor frequência na tabela dada neste exemplo é 2, para “Mulher”-“Sim”. Vamos, então, trabalhar essa primeira célula. Podemos construir outras duas tabelas com os mesmos totais marginais (em negrito, nas tabelas), com resultados 1 e zero, para “Mulher”-“Sim”e calcular respectivas probabilidades de ocorrência. Veja a distribuição dos dados, para uma mulher dizendo “Sim”:

Pacientes classificados segundo o sexo e a queixa de dor (Tabela auxiliar)

Sexo	**Queixa de dor**		**Total**
	Sim	**Não**	
Mulher	1	4	**5**
Homem	7	3	**10**
Total	**8**	**7**	**15**

A probabilidade de uma mulher dizer “Sim” é:

$$P(1,S) = \frac{(1+4)!\,(7+3)!\,(1+7)!\,(4+3)!}{15!\,1!\,4!\,7!\,3!} = 0{,}093$$

Veja a distribuição dos dados para nenhuma (zero) mulher dizendo “sim”:

Pacientes classificados segundo o sexo e a queixa de dor (Tabela auxiliar)

Sexo	**Queixa de dor**		**Total**
	Sim	**Não**	
Mulher	0	5	**5**
Homem	8	2	**10**
Total	**8**	**7**	**15**

A probabilidade de ocorrerem esses valores é:

$$P(0,S) = \frac{(0+5)!\,(8+2)!\,(0+8)!\,(5+2)!}{15!\,0!\,5!\,8!\,2!} = 0{,}007$$

Então, a probabilidade de ocorrerem valores iguais ou mais extremos (zero ou 1) do que 2 mulheres dizendo “Sim” (para um teste unilateral) é:

$$P(2,S) + P(1,S) + P(0,S) = 0,\ 326 + 0{,}093 + 0{,}007 = 0{,}426$$

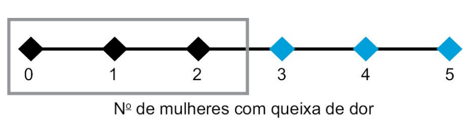


Figura 5.7 – Valores correspondentes ao p-valor para teste unilateral.

Como a hipótese alternativa é a de que a proporção de pessoas que se queixam de dor é diferente nos dois sexos, o teste é *bilateral*. Calculamos a probabilidade de 2 ou menos mulheres *terem se queixado* de dor. Vamos agora para o outro lado, isto é, calcular as probabilidades de 3, 2, 1 e 0 mulheres *não terem se queixado* de dor.

Veja: 3 mulheres não terem se queixado de dor foi o valor observado. Essa probabilidade já foi calculada: é 0,326. Vamos, então, calcular as probabilidades de ocorrerem 2, 1 e 0 mulheres dizendo “Não”.

$$P(2,N) = \frac{(3+2)!\,(5+5)!\,(3+5)!\,(2+5)!}{15!\,3!\,2!\,5!\,5!} = 0{,}392$$

$$P(1,N) = \frac{(4+1)!\,(4+6)!\,(4+4)!\,(1+6)!}{15!\,4!\,1!\,4!\,6!} = 0{,}163$$

$$P(0,N) = \frac{(5+0)!\,(3+7)!\,(5+3)!\,(0+7)!}{15!\,5!\,0!\,3!\,7!} = 0{,}019$$

Olhe as probabilidades calculadas e constate: a probabilidade de duas mulheres dizendo “Não” é maior do que a probabilidade de ocorrência dos valores observados. Então, esse *não* é um valor extremo.

Quarto passo: Some as probabilidades calculadas, de ocorrerem valores iguais ou mais extremos que os observados, dos dois lados:

$$P(2,S) + P(1,S) + P(0,S) + P(1,N) + P(0,N) =$$

$$0,326 + 0{,}0093 + 0{,}007 + 0{,}163 + 0{,}019 = 0{,}608$$

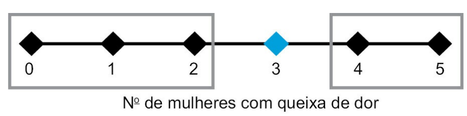


Figura 5.8 – Valores correspondentes ao *p*-valor para teste bilateral.

Quinto passo: A probabilidade calculada é o *p*-valor. Então, você não rejeita a hipótese de que a proporção de pessoas que se queixam de dor é a mesma nos dois sexos.

5.4. TESTE DE MCNEMAR

O teste de McNemar testa a *consistência das respostas de dois grupos*, desde que as respostas dos participantes sejam binárias, isto é, do tipo "sim" ou "não". Portanto, o teste de McNemar *não* testa a independência de duas variáveis. Por exemplo, para estudar o efeito de uma intervenção, o pesquisador precisa de duas amostras de pessoas: as que respondem "sim" e as que respondem "não". Faz a intervenção e estuda a consistência das respostas nos dois grupos, ou seja, se nos dois grupos as mudanças ocorreram na mesma proporção. Veja a Figura 5.9.

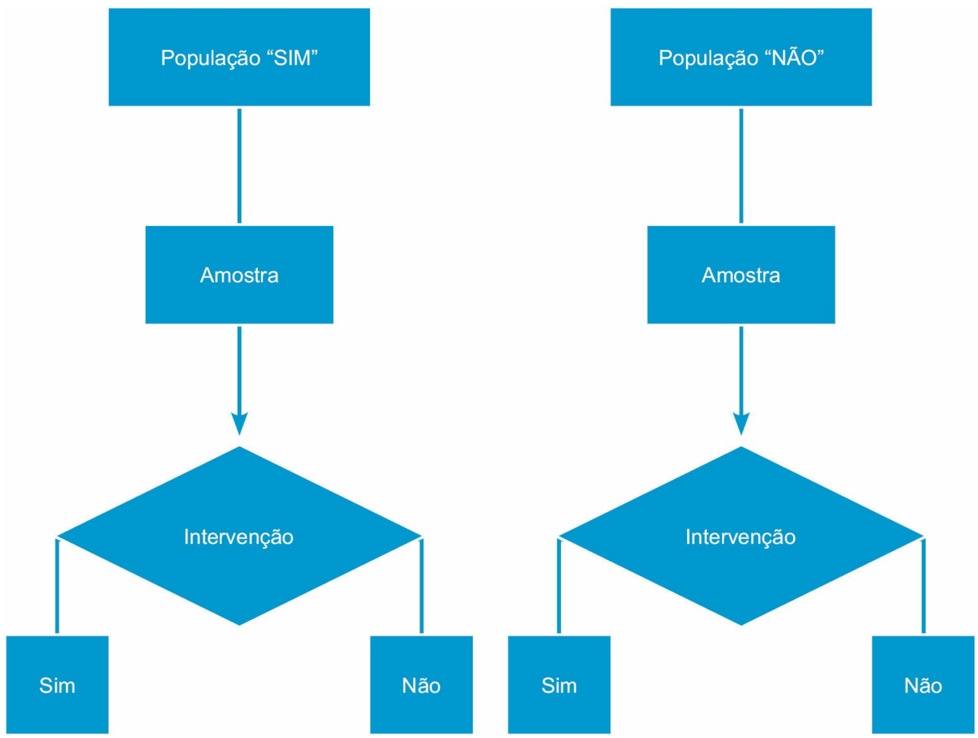


Figura 5.9 – Consistência de respostas.

Para fazer o teste:

Primeiro passo: A hipótese da nulidade é a de que o número de pessoas que mantêm a resposta inicial é a mesma nos dois grupos. A hipótese alternativa é a de que esses números são diferentes. Estabeleça o nível de significância.

Segundo passo: Organize as respostas em uma tabela 2 × 2, de tal maneira que as pessoas que mudaram a resposta (diziam "Sim" e passaram a dizer "Não" ou diziam "Não" e passaram a "Sim") sejam indicados por *a* e *d*. Calcule a estatística de teste:

$$\chi^2 = \frac{(a-d)^2}{a+d} \tag{5.6}$$

que tem distribuição de χ2 com 1 grau de liberdade.

Terceiro passo: Compare o valor calculado de χ2 que está associado a 1 grau de liberdade com o valor crítico de χ2 no nível estabelecido de significância (Tabela 2 do Apêndice). Rejeite a hipótese da nulidade toda vez que o valor calculado de χ2 for maior do que o valor crítico. Se você estiver usando um programa de computador para fazer esses cálculos, vai encontrar o *p*-valor.

Exemplo 5.13

Imagine um ensaio para estudar efeitos adversos de um fármaco indicado para casos de hipertensão. Os 104 pacientes, com o mesmo perfil social e demográfico, recém-diagnosticados e sem tratamento, responderam a diversas questões, tanto no início do tratamento como 12 semanas após. No início, 32 pacientes queixavam-se de cefaleia e 72 não tinham essa queixa.

No final do tratamento, 19 dos 32 que se queixavam antes *não* tinham mais a queixa e, dos 72 que *não* tinham queixa de cefaleia, 1 passou a se queixar.

Para fazer o teste de χ2 de McNemar:

Primeiro passo: A hipótese da nulidade é a de que o tratamento não tem efeito sobre as queixas de cefaleia. Sob essa hipótese, espera-se que a probabilidade de que um paciente mude de situação seja a mesma nos dois grupos. Seja α = 0,05.

Segundo passo: Organize a tabela 2 × 2 de tal maneira que as pessoas que mudaram a resposta sejam indicadas por *a* e *d*. Veja que:

20 pacientes mudaram de resposta: 19 queixavam-se de cefaleia “antes” do tratamento e não se queixavam “depois”. Então, *a* = 19. Um paciente não se queixava “antes”, mas passou a se queixar “depois”. Então, *d* = 1.

Pacientes com queixas de cefaleia antes e depois de tratados

Antes do tratamento	**Depois do tratamento**	
	Sem queixa	**Com queixa**
Sem queixa	19	13
Com queixa	71	1

Fonte: Dados fictícios com base em LUNA, R.L. Eficácia e tolerabilidade da associação bisoprolol/ hidroclorotiazida na hipertensão arterial. *Arq Bras Cardiol* 71(4): 601-608, 1998.

Segundo passo: Calcule o χ2 de McNemar:

$$\chi^2 = \frac{(19-1)^2}{19+1} = \frac{18^2}{20} = 16{,}20$$

Terceiro passo: Como o valor calculado de χ2 (16,20) com 1 grau de liberdade é maior do que o valor crítico no nível de 5% de significância (3,84), o resultado é significante. As queixas de cefaleia diminuíram com o tratamento.

Se você estiver usando um programa de computador para fazer esses cálculos, vai encontrar *p*-valor = 0,000144. Como o *p*-valor é menor do que 0,05 ($p < 0{,}05$), a conclusão é a de que as queixas de cefaleia diminuíram com o tratamento

5.4.1.Correção de continuidade

Muitos estatísticos recomendam calcular o χ2 de McNemar *com correção de continuidade*,

principalmente quando a amostra é pequena. Os programas de computador em geral apresentam essa correção. A fórmula (5.6) fica, então, como segue:

$$\chi^2 = \frac{(|a-d|-1)^2}{a+d} \tag{5.7}$$

A estatística de teste com correção de continuidade é menor do que o valor calculado sem a correção. Então, a correção de continuidade produz um teste mais *conservador*, que tem menor probabilidade de rejeitar a hipótese da nulidade. Se a amostra for pequena, o efeito da correção de continuidade será ainda maior.

Exemplo 5.14

Reveja o exemplo 5.12. Como os programas de computador em geral fazem a correção de continuidade, verifique:

$$\chi^2 = \frac{(|19-1|-1)^2}{19+1} = \frac{17^2}{20} = 14,45$$

O valor calculado de χ2 (14,45) é maior do que o valor crítico com 1 grau de liberdade e no nível de 5% de significância (3,84). Então o resultado é significante no nível de 5%.

Se você usar um programa para computador, encontrará *p*-valor = 0,0001.

Observação

Se as frequências observadas forem menores do que 5, não calcule o valor do χ2 de McNemar; faça outro teste.[77]

RESUMO E OBJETIVO DO CAPÍTULO

A tabelas de contingência 2 × 2 são muito frequentes nas publicações da área da saúde. Depois de ler este Capítulo, você será capaz de, diante uma tabela 2 × 2, saber escolher, aplicar e indicar um dos testes:

1. teste de χ2 de Pearson;
2. teste de proporções;
3. teste exato de Fisher;
4. teste de χ2 de McNemar.

Saberá, ainda, que a correção de continuidade produz um teste mais conservador, mas é indicada se a amostra for pequena.

---

[66] É usual apresentar o teste de qui-quadrado em textos que tratam de testes não paramétricos. No entanto, muitos estatísticos consideram que, pelo fato de comparar proporções, o teste de qui-quadrado não seria paramétrico. Veja Daniel, W.W. *Applied nonparametric statistics.* 2nd ed. Pacific Grove: Duxbury, 1990. p. 178.

[67] Em inglês, a letra χ é escrita "chi" e se lê "cai". Em português, escreva qui.

[68] As restrições foram estabelecidas por COCHRAN, W.O. Some methods for strengthening the common χ^2 tests. *Biometrics* 10: 417-451, 1954.

[69] Existe quem argumente contra a correção de continuidade. Veja: GRIZZLE, J.E. Continuity correction in the χ2 test for 2 x 2 tables. *The American Statistician* 21: 28-32, 1967.

70 O teste de χ^2 leva ao mesmo resultado. Veja Teste de χ2 em soniavieira.blogspot.com.br.

[71] Fonte: SCARPELINl, S.; ANDRADE, J.I., STRACIERE, L.D.S.; GRADE, M.H.C.; MACCHETI, A.H., PASSOS, A.D.C. Estudo comparativo entre o tratamento não operatório e a cirurgia conservadora no trauma esplênico. *Rev Col Bras Cir* 26(5): 281-284, 1999.

[72] No Exemplo 5.9, a distribuição é apenas aproximadamente normal, porque $np < 5$, mas a amostra é grande.

[73] Os valores de χ^2 com um grau de liberdade são iguais aos valores do quadrado de z (da distribuição normal reduzida).

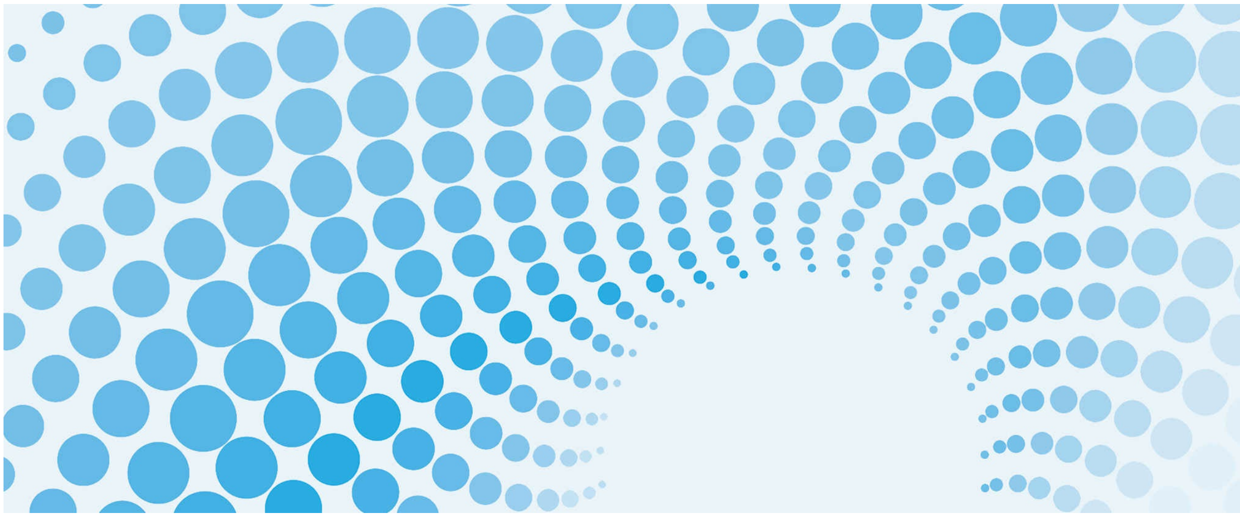
[74] Veja, por exemplo, FLEISS, J.L. *Statistical methods for rates and proportions.* 2nd ed. Nova York: Wiley, 1981. p. 22-24.

[75] O teste se baseia na distribuição hipergeométrica.

[76] Os dados são de

PREACHER, KJ. Disponível em: Interactive Fisher's Exact Test – Quantpsy.org. quantpsy.org/fisher/fisher.htm. Acesso em 4 de novembro de 2017.

[77] SIEGEL, S. *Nonparametrics Statistics for the behavioral sciences*. 2nd ed. Nova York: McGraw-Hill, 1956. Note que o exemplo não seguiu esta recomendação.



5.5.1 Calcule os valores de χ2 com e sem a correção de continuidade para os dados apresentados no Exemplo 5.1. Verifique que a correção de continuidade faz diminuir o valor da estatística.

5.5.2 Das 300 pessoas que chegavam de um voo a um grande aeroporto,[78] 81 admitiram ter medo de voar e, das 200 pessoas que embarcavam em outro voo, mas no mesmo dia e na mesma hora, 32 admitiram também ter medo. Calcule as proporções (medo na chegada e medo na partida) e faça um teste de χ2.

5.5.3 Os dados apresentados na Tabela 5.4 foram obtidos por meio de estudo transversal com 1.330 nipo-brasileiros de primeira e segunda gerações, de ambos os sexos, com mais de 30 anos. Com base no índice de massa corporal, os nipo-brasileiros de primeira e segunda gerações foram classificados como tendo ou não sobrepeso e obesidade. Faça o teste de χ2 e dê a interpretação tendo em vista a amostra estudada.

Tabela 5.4 – Nipo-brasileiros com sobrepeso segundo a geração (Bauru SP, 2000)

Geração	**Sobrepeso**	
	Sim	**Não**
Primeira	82	175
Segunda	525	548

Fonte: SIMONY, R.F. et al. Prevalência de sobrepeso e obesidade em nipo-brasileiros: comparação entre sexos e geração. *Rev Nutr* 21(2): 169-176, 2008.

5.5.4. Foi feito um ensaio clínico para estudar o efeito da betametasona sobre a dor no pós-operatório de pacientes submetidos a tratamento endodôntico. O experimento era duplo-cego, randomizado e foram administrados dois comprimidos de betametasona para o grupo experimental e dois comprimidos de placebo para o grupo controle antes da instrumentação. Os dados estão na Tabela 5.5. Faça o teste.

Tabela 5.5 – Participantes da pesquisa segundo o grupo e o relato de dor

Grupo	**Relato de dor**	
	Sim	**Não**
Controle	9	12
Experimental	15	2

Fonte: QUINTANA-GOMES, J.R.V.; ANDRADE, E.K. Estudo clínico dos efeitos da betametasona sobre a incidência da dor após a instrumentação endodôntica. J Bras Odont Clinica 2(12): 73-76. [s.d.]

5.5.5. Acredita-se que certo tratamento prolongue a vida de pessoas que sofreram um ataque cardíaco.[79] Foi feito um ensaio clínico em que cinco pacientes receberam o novo tratamento (grupo tratado) e cinco receberam o tratamento convencional (grupo controle). Os 10 pacientes foram cuidadosamente selecionados: mesma faixa de idade, mesma gravidade do ataque, mesmo estado geral de saúde, sendo designados aos grupos por processo aleatório. Cinco anos depois, estavam vivos 4 pacientes do grupo tratado e 2 do grupo controle. Existe evidência de que o tratamento ajudou?

5.5.6. Para verificar se doença periodontal (gengivite) está associada ao hábito de fumar, realizou-se um estudo envolvendo policiais militares. Os participantes da pesquisa foram classificados em dois grupos: casos (com doença periodontal) e controles (sem doença periodontal). Todos estavam na mesma faixa de idade, com o mesmo estado civil, na mesma patente militar e no mesmo nível socioeconômico. Perguntou-se aos policiais militares se eles fumavam ou não. Os dados estão na Tabela 5.6. Faça o teste.

Tabela 5.6 – Hábito de fumar e gengivite

Grupo	Hábito de fumar	
	Sim	**Não**
Casos	26	15
Controles	25	29

Fonte: CRUZ, C.M.N. *Estresse e fumo como fatores de risco para a doença periodontal.* Tese de mestrado. Campinas: Unicastelo, 2000.

5.5.7. Imagine um ensaio clínico com 1.455 participantes: 1.165 tinham dor nos joelhos e 290 não tinham dor nos joelhos. Eles receberam um tratamento em teste. Depois de tratados, 215 dos que não tinham dor nos joelhos continuaram não tendo e 380 dos que tinham dor nos joelhos continuaram a ter. Importante: note que os números são exatamente iguais aos da tabela do Exemplo 5.6, onde fizemos o teste de qui-quadrado de Pearson. Aqui, faça o teste de qui-quadrado de McNemar e explique o porquê.

Tabela 5.7 – Participantes da pesquisa segundo a dor antes e depois de tratados

Tinham dor antes	Tinham dor depois		Total
	Não	**Sim**	
Não	215	75	290
Sim	785	380	1165
Total	1000	455	1455

Fonte: http://www.theanalysisfactor.com/difference-between-chi-square-test-and-mcnemar-test/

5.5.8. Para verificar[80] se o distúrbio da articulação temporomandibular (DTM) em crianças e adolescentes está associado ao sexo, foram examinados 604 estudantes com idades entre 8 e 17 anos, moradores de Piracicaba, SP. Foi constatada a disfunção em 118 dos 264 estudantes do sexo masculino e em 117 estudantes do sexo feminino. Calcule o valor de qui-quadrado e interprete.

5.5.9. No ensaio controlado e randomizado denominado Women's Health Initiative (Iniciativa para a Saúde das Mulheres),[81] 16.608 mulheres em menopausa com idades entre 50 e 79 anos foram recrutadas em clínicas dos Estados Unidos entre 1993 e 1998 para receberem hormônio (estrógeno e progesterona) ou placebo, com a finalidade de levantar os riscos e os benefícios associados ao uso dessa combinação de hormônios. No ensaio, 8.506 mulheres receberam o hormônio e 8.102 receberam placebo. Na avaliação das características dessas mulheres após a randomização, verificou-se que, no grupo tratado, 1.623 usavam aspirina e, no grupo controle, 1.631. Teste a hipótese de que a proporção de usuárias de aspirina é a mesma nos dois grupos.

5.5.10. Os traumas faciais ocorrem, predominantemente, em adultos jovens, mas houve um aumento de incidência entre idosos. Os dados[82] apresentados na Tabela 5.8 buscam verificar se existe associação entre sexo e etiologia da fratura (queda e outras causas, dentre as quais estão agressão e acidentes de trânsito). Faça o teste.

Tabela 5.8 – Distribuição dos participantes da pesquisa segundo o sexo e a etiologia da fratura na face

Sexo	Etiologia		Total
	Queda	**Outras causas**	
Homens	11	15	26
Mulheres	12	5	17
Total	23	20	43

[78] FREUND, J.E.; SMITH, R.M. *Statistics: a first course*. 4th ed. Englewood Cliffs: Prentice Hall, 1970. p. 411.

[79] HODGES, J.L., LEHMAN, E.L. *Basic Concepts of Probability and Statistics*. 2nd ed. San Francisco: Holden Day, 1970.

[80] COSTA, L.F.; GUIMARÃES, J.P.; CHAOBAH, A. Prevalência de distúrbios da articulação temporomandibular em crianças e adolescentes brasileiros e sua relação com má oclusão e hábitos parafuncionais: um estudo epidemiológico transversal. *J Bras Ortodon Ortop Facial 9*(49): 67-74, 2004.

[81] Women's Health Initiative Randomized Controlled Trial. Risks and benefits of estrogen plus progestin in healthy post menopausal women. *JAMA* 288 (3): 321-333, 2002.

[82] Chrcanovic, B.R.; Souza, L.N.; Freire-Maia, B. Fraturas de face em idosos: estudo retrospectivo de um ano em hospital público de Belo Horizonte, MG. *Rev ABO Nac* 16(1): 39-44, 2008.

6

Tabelas $2 \times s$

Neste Capítulo são examinadas algumas estatísticas que estudam a relação entre duas variáveis categorizadas, apresentadas em *tabelas de contingência* $2 \times s$, em que s indica o número de colunas.

Exemplo 6.1

Imagine que, para verificar se sexo e estado antropométrico estão associados, pesquisadores classificaram 1.000 universitários segundo essas duas variáveis e apresentaram os dados em uma tabela.

Participantes da pesquisa segundo o sexo e o estado antropométrico

Sexo	**Estado antropométrico**		
	Magreza	**Eutrofia**	**Sobrepeso**
Masculino	200 (20%)	150 (15%)	50 (5%)
Feminino	250 (25%)	300 (30%)	50 (5%)

Você tem, neste exemplo, uma tabela $2 \; {}^{3}_{2} \; 3$ para estudar a relação entre sexo e estado antropométrico.

6.1. TESTE DE χ2 DE PEARSON

Indicação: O teste de χ2 de Pearson é indicado para testar a hipótese de que *duas variáveis qualitativas são independentes.* Veja o *procedimento.*

Primeiro passo: Estabeleça as hipóteses e o nível de significância.

- H_0: as duas variáveis são independentes.
- H_1: as duas variáveis *não* são independentes.

Nível de significância: normalmente são estabelecidos níveis de significância de 0,01 ou 0,05 ou 0,10.

Segundo passo: Calcule as frequências esperadas sob H_0. Para obter a *frequência esperada de* determinada célula:

1. Multiplique o *total marginal de cada coluna* pelo *total marginal da respectiva linha*.
2. Divida o resultado pelo *tamanho da amostra*.
3. Verifique: os *totais* das frequências esperadas são, obrigatoriamente, iguais aos das frequências observadas.

Terceiro passo: Calcule a estatística de teste.

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{6.1}$$

Na fórmula, *O* indica frequências observadas e *E* indica frequências esperadas.

Quarto passo: O valor calculado de χ2 está associado a (2 – 1) (*s* – 1) graus de liberdade. Compare o valor calculado de χ2 com o valor crítico (Tabela 2 do Apêndice) no nível de significância estabelecido. Toda vez que o valor calculado da estatística for igual ou maior do que o valor crítico, rejeite a hipótese da nulidade. Se você estiver usando um programa para computador, obterá o *p*-valor.

Exemplo 6.2

Reveja o exemplo 6.1: os dados são as *frequências observadas* pelos pesquisadores. Calcule os *totais marginais* e o *total geral*.

Participantes da pesquisa segundo o sexo e o estado antropométrico

Sexo	**Estado antropométrico**			**Total**
	Magreza	**Eutrofia**	**Sobrepeso**	
Masculino	200	150	50	400
Feminino	250	300	50	600
Total	450	450	100	1.000

Para fazer o teste de $\chi2$:

Primeiro passo: A hipótese da nulidade estabelece que o estado antropométrico não depende do sexo; a hipótese alternativa estabelece que sexo e estado antropométrico são dependentes. Nível de significância: 5%.

Segundo passo: Calcule as frequências esperadas sob H_0. Assim, para obter a *frequência esperada de magros de sexo masculino*:

1. Multiplique o total de magros (450) pelo total de pessoas do sexo masculino (400).
2. Divida o resultado pelo tamanho da amostra (1.000).

Para achar as demais frequências esperadas, proceda de maneira análoga. Veja os cálculos na tabela.

Cálculo das frequências esperadas

Sexo	**Estado antropométrico**			**Total**
	Magreza	**Eutrofia**	**Sobrepeso**	
Masculino	450 × 400 / 1.000	450 × 400 / 1.000	100 × 400 / 1.000	400
Feminino	450 × 400 / 1.000	450 × 600 / 1.000	100 × 600 / 1.000	600
Total	450	450	100	1.000

Fazendo os cálculos, você obtém as frequências esperadas:

Frequências esperadas de participantes da pesquisa segundo o sexo e o estado antropométrico

Sexo	**Estado antropométrico**			**Total**
	Magreza	**Eutrofia**	**Sobrepeso**	
Masculino	180	180	40	400
Feminino	270	270	60	600
Total	450	450	100	1.000

Terceiro passo: Calcule a estatística de teste dada pela fórmula (6.1). Para fazer os cálculos:

1. Calcule as diferenças entre os valores observados e os valores esperados.
2. Eleve as diferenças ao quadrado.
3. Divida os quadrados das diferenças pelos respectivos valores esperados.
4. Some.

Para obter o valor de $\chi2$, são necessários diversos cálculos.

Cálculos auxiliares para obtenção do $\chi2$

Sexo	**Estado antropométrico**	***O***	***E***	***O-E***	**$(O\text{-}E)^2$**	**$(O\text{-}E)^2/E$**
Masculino	Magreza	200	180	20	400	2,222
	Eutrofia	150	180	–30	900	5,000
	Sobrepeso	50	40	10	100	2,500
Feminino	Magreza	250	270	–20	400	1,481
	Eutrofia	300	270	30	900	3,333
	Sobrepeso	50	60	–10	100	1,667
Soma		1.000	1.000	0		16,204

Quarto passo: O valor calculado de χ2 é 16,204, com (2 – 1) (3 – 1) = 2 graus de liberdade. No nível de 5% de significância e com 2 graus de liberdade, o valor crítico de χ2 (Tabela 2 do Apêndice) é 5,99. Como o valor calculado de χ2 é maior do que o valor crítico, rejeite a hipótese de independência. O estado antropométrico depende do sexo. Se você estiver usando um programa para computador, obterá *p*-valor = 0,000304.

6.1.1. Exigências teóricas para aplicação do teste de χ2

Preste atenção às exigências:

1. Independência: as características em comparação devem ser independentes, como, por exemplo, um grupo ser constituído por homens e o outro, por mulheres.
2. Frequências esperadas: se $s > 2$, menos de 20% das frequências esperadas podem ser menores do que 5. Nenhuma frequência esperada pode ser inferior a 1. Se isso acontecer, junte categorias.

IMPORTANTE:

A tabela 2 × 2 vista no Capítulo 5 é um caso particular da tabela de contingência 2 × *s* vista aqui. Então, a fórmula para o cálculo de χ2 apresentada neste Capítulo pode ser usada para tabelas 2 × 2. Os resultados serão iguais porque as fórmulas são iguais, como se pode demonstrar por simples álgebra. Aquela é mais fácil de aplicar e revelou a vantagem adicional de usar a mesma notação usada posteriormente, em outros testes apresentados naquele Capítulo.

6.1.2. Algumas questões teóricas

Você pode deixar de ler esta seção sem prejuízo na leitura do restante do livro, mas talvez tenha interesse em obter mais algumas informações sobre o teste de χ2 .

6.1.3. Graus de liberdade

O que significa grau de liberdade? É o número de observações que podem variar livremente – ou seja, é o número de observações “livres para variar”. Nas tabelas de contingência, o total das linhas e o total das colunas são “fixos” – quando a amostra já foi coletada.

Para tornar a explicação mais concreta, veja a tabela 2 × 2 dada em seguida: os totais das linhas e das colunas são “fixos” porque, terminada a coleta de dados, você sabe que foram entrevistados 200 homens e 200 mulheres. Houve 300 respostas “Sim” e 100 respostas “Não”.

Antes de tabular os dados, alguém poderia imaginar, por exemplo, que 150 mulheres disseram “Sim”. Se isso fosse verdade, automaticamente, os demais valores estariam determinados. Só há liberdade para “fazer variar” um valor. Então, o *grau de liberdade* é 1.

Sexo	Opinião		Total
	Sim	Não	
Mulheres	150		200

Homens			200
Total	300	100	400

Veja agora a tabela 2 × *s*. As células em branco (poderiam ser quaisquer outras) estão livres para variar. As células com pontos ficam, então, determinadas porque os totais são “fixos”. Os graus de liberdade são, portanto, (2 – 1) (*s* – 1).

Grupos		**Grau da doença**		
	Leve	**Moderado**	**Severo**	**Total**
A			•	100
B	•	•	•	100
Total	200	150	50	400

6.1.4. Frequências esperadas

Volte ao Exemplo 6.1: Os dados são as *frequências observadas* pelos pesquisadores. Para aplicar o teste de χ2:

Primeiro passo: Estabeleça a hipótese da nulidade, de que o estado antropométrico não depende do sexo. Estabeleça a hipótese alternativa, de que o estado antropométrico está associado ao sexo. Estabeleça o nível de significância.

Segundo passo: Considere verdadeira a hipótese da nulidade de que o estado antropométrico não depende do sexo. Sob essa hipótese:

• A proporção de magros deve ser igual nos dois sexos.
• A proporção de eutróficos deve ser igual nos dois sexos.
• A proporção de adolescentes com sobrepeso e obesos deve ser igual nos dois sexos.

Calcule as *proporções esperadas* sob a hipótese da nulidade. Se são 450 magros em uma amostra de 1.000 participantes, *sob* H_0 a *proporção esperada de magros* é

$$p = \frac{total\ de\ magros}{tamanho\ da\ amostra} = \frac{450}{1.000} = 0,45$$

São 450 eutróficos em uma amostra de 1.000 participantes. Logo, *sob* H_0 a *proporção esperada de eutróficos* é

$$p = \frac{total\ de\ eutróficos}{tamanho\ da\ amostra} = \frac{450}{1.000} = 0,45$$

São 100 participantes com sobrepeso em uma amostra de 1.000 participantes. Logo, sob H_0, a *proporção esperada de sobrepesados* é:

$$p = \frac{total\ de\ sobrepesados}{tamanho\ da\ amostra} = \frac{100}{1.000} = 0{,}10$$

Foram observados mais participantes de sexo feminino (600) do que do sexo masculino (400). Para obter a *frequência esperada* de cada estado antropométrico em cada sexo, multiplique a proporção que se espera de cada estado antropométrico sob H_0 pelo total de participantes desse sexo. Veja os resultados na tabela.

Cálculo das frequências esperadas

Sexo	**Estado antropométrico**			**Total**
	Magreza	**Eutrofia**	**Sobrepeso**	
Masculino	0,45 × 400	0,45 × 400	0,10 × 400	400
Feminino	0,45 × 600	0,45 × 600	0,10 × 600	600
Total	450	450	100	1.000

Frequências esperadas de participantes da pesquisa segundo o sexo e o estado antropométrico

Sexo	**Estado antropométrico**			**Total**
	Magreza	**Eutrofia**	**Sobrepeso**	
Masculino	180	180	40	400
Feminino	270	270	60	600
Total	450	450	100	1.000

6.2. PARTIÇÃO DAS TABELAS 2 × S

O teste de $\chi 2$ tem interpretação óbvia quando é aplicado a uma tabela 2 × 2. No entanto, se uma das variáveis tiver mais de duas categorias, isto é, se os dados estiverem apresentados em uma tabela de contingência 2 × *s*, a interpretação *não* é direta.

Isso porque, se você rejeitar a hipótese de independência, fica sem saber se a associação ocorre entre *todas as categorias* de uma variável ou apenas entre *certas categorias* dessa variável.

É possível fazer uma partição da tabela original. Em outras palavras, é possível *quebrar uma tabela grande* em outras menores – e analisar essas tabelas menores separadamente. Recomenda-se, porém, que o número de partições seja igual ou menor do que o número de graus de liberdade do teste inicial.

Mesmo assim, faça essa partição somente se

1) Tiver alguma hipótese formulada com base na literatura ou
2) A análise inicial tiver fornecido evidência contra a hipótese de independência (você será mais convincente se tiver usado nível de significância de 1%).

Exemplo 6.3

Reveja o Exemplo 6.1: para verificar se sexo e estado antropométrico estão associados, pesquisadores classificaram 1.000 universitários segundo essas duas variáveis. O *p*-valor do teste de $\chi 2$ aplicado aos dados é 0,000304. Pode, portanto, ser feita a partição da tabela.

Vamos, primeiramente, aplicar o teste de $\chi 2$ para verificar se o sexo está associado ao peso adequado. Para fazer isso, vamos somar participantes com peso inadequado (magreza e sobrepeso) de cada sexo para comparar com os de peso adequado.

Participantes da pesquisa segundo o sexo e o peso (adequado ou não)(proporções entre parênteses)

Sexo	**Peso adequado**		**Total**
	Sim	**Não**	
Masculino	150 (15%)	250 (25%)	400
Feminino	300 (30%)	300 (30%)	600
Total	450	550	1.000

O resultado do teste de $\chi 2$ aplicado a essa tabela é 15,152 e o *p*-valor é 0,0001. A conclusão é de que o sexo está associado ao estado antropométrico: pesos inadequados são mais comuns no sexo masculino. Vamos estudar agora a associação entre sexo e magreza e sobrepeso.

Participantes da pesquisa com peso inadequado classificado, segundo o sexo e o estado antropométrico (proporções entre parênteses)

Sexo	**Estado antropométrico**		**Total**
	Magreza	**Sobrepeso**	

Masculino	200 (36,36%)	50 (9,09%)	250
Feminino	250 (45,45%)	50 (9,09%)	300
Total	450	100	550

O resultado do teste de $\chi 2$ aplicado a essa tabela é 1,019 e o *p*-valor é 0,3129. Não se rejeita a hipótese de independência. Magreza e sobrepeso aparecem igualmente nos dois sexos.

É difícil interpretar a associação em tabelas de contingência quando uma das variáveis tem mais de duas categorias. Por essa razão, vamos ver outro exemplo.

Exemplo 6.4

Para comparar uma nova droga (grupo tratado) com uma droga conhecida (controle positivo) na solução da calvície, um pesquisador acha importante ter um grupo usando placebo (controle negativo). Foram, então, colocadas duas questões: primeira, "droga tem efeito?" e segunda, "a nova droga tem efeito diferente do efeito da droga conhecida?". O efeito da droga (positivo ou nenhum) sobre a calvície foi avaliado pelos próprios participantes da pesquisa.

Participantes da pesquisa segundo o grupo e o fato de terem sentido efeito positivo ou não

Grupo	**Efeito positivo**		**Total**
	Sim	**Não**	
Tratado	64	36	100
Controle positivo	60	40	100
Controle negativo	26	74	100
Total	150	150	300

Foi aplicado o teste de $\chi 2$ de Pearson. O valor calculado de $\chi 2$ com $(2 – 1)\ (3 – 1) = 2$ graus de liberdade é 34,88, significante no nível de 1%. A conclusão é de que existe diferença entre grupos. Qual grupo teve melhor resultado? Para responder a essa questão, vamos fazer a *partição da tabela*.

Para isso, somamos os dados dos participantes que receberam a nova droga com os dados dos participantes que receberam a droga conhecida para constituir um novo grupo – o grupo tratado. Esse novo grupo é comparado com o controle negativo (placebo) por meio de um teste de $\chi 2$.

Participantes da pesquisa segundo o grupo e o fato de terem sentido efeito positivo

Grupo	**Efeito positivo**			**Proporção**
	Sim	**Não**	**Total**	**Sim**
Tratado	124	76	200	0,620
Controle negativo	26	74	100	0,260
Total	150	150	300	0,500

O valor calculado de $\chi 2$ com 1 grau de liberdade é 34,56, significante no nível de 1% (Tabela 2 do Apêndice). A proporção de participantes da pesquisa que registrou efeito positivo foi maior no grupo tratado. No entanto, qual é a melhor droga, a nova ou a convencional? Compare o grupo tratado (nova droga) com o grupo controle positivo (droga convencional) por meio de um teste de $\chi 2$.

Participantes da pesquisa segundo o grupo e o fato de terem sentido efeito positivo

Grupo	Efeito positivo			Proporção
	Sim	**Não**	**Total**	**Sim**
Tratado	64	36	100	0,640
Controle positivo	60	40	100	0,600
Total	124	76	200	0,620

O valor calculado de $\chi 2$ com 1 grau de liberdade é 0,34, não significante no nível de 5%. Então, não foi detectada diferença significante entre as drogas.[83]

6.3. PROCEDIMENTO DE MARASCUILO

Indicação: O procedimento de Marascuilo faz *comparações múltiplas*, isto é, compara proporções populacionais, duas a duas.

Se você aplicou o teste de χ2 a uma tabela de contingência 2 × *s* e obteve resultado significante, pode fazer a partição da tabela. Isso nem sempre resolve seu problema, porque o número de partições deve ser igual ou menor ao número de graus de liberdade do teste. Na verdade, você quer um teste de comparações múltiplas, mas não existe uma técnica estatística exata para comparações de proporções duas a duas. Você pode, porém, optar pelo procedimento de Marascuilo,[84] que não é tão conhecido, mas faz comparações múltiplas.

Exemplo 6.5

Foi feito um estudo multicêntrico para testar o efeito de um anti-hipertensivo sobre derrame recorrente. O protocolo permitia que os diferentes centros médicos envolvidos na pesquisa administrassem, como coadjuvante, o tratamento que sempre haviam indicado. Um pesquisador suspeitou de que esses tratamentos coadjuvantes poderiam ter efeito sobre a ocorrência de novo derrame. Para testar essa hipótese, o pesquisador resolveu analisar novamente os dados apresentados na tabela a seguir. Aplicou o teste de χ2.

Participantes da pesquisa segundo o centro em que foram tratados e o fato de terem ou não tido derrame recorrente

Derrame recorrente	**Centro**				**Total**
	A	**B**	**C**	**D**	
Sim	16	12	21	12	61
Não	179	70	78	54	381
Total	195	82	99	66	442

O valor calculado de χ2 está associado a (2 – 1) (4 –1) = 3 graus de liberdade. É 10,7603, significante no nível de 5%. A recorrência de derrame depende, portanto, do centro (e provavelmente do tratamento prescrito nesse centro) em que o paciente é tratado. É correto o pesquisador perguntar: em quais centros a probabilidade de derrames recorrentes é significantemente maior? Aplique o procedimento de Marascuilo.

Primeiro passo: Estabeleça a hipótese da nulidade (H_0), de que as proporções são iguais em todas as categorias. Estabeleça a hipótese alternativa (H_1), de que existe pelo menos uma proporção diferente das demais. Estabeleça o nível de 5% de significância.

Segundo passo: Calcule as proporções nas diversas categorias.

Terceiro passo: Calcule as amplitudes críticas para o procedimento de Marascuilo:

$$d_{ij} = \sqrt{\chi^2_{(s-1)}} \sqrt{\frac{p_i(1-p_i)}{n_i} + \frac{p_j(1-p_j)}{n_j}} \qquad (6.2)$$

Nessa fórmula:

$$i = 1, 2,...,s;$$

$$j = 1, 2,...,s;\ i \neq j$$

$\chi^2_{(s-1)}$ é o valor de $\chi2$ com (s – 1) graus de liberdade, no nível de significância considerado.

Quarto passo: Calcule as diferenças entre as proporções observadas duas a duas. Compare cada diferença calculada com a respectiva amplitude crítica. Uma diferença é estatisticamente significante se exceder a amplitude crítica.

Exemplo 6.6

O Exemplo 6.5 apresentou os dados de um estudo multicêntrico para testar o efeito de um anti-hipertensivo sobre derrame recorrente. O valor calculado de $\chi2$, com 3 graus de liberdade, é significante no nível de 5%. A recorrência de derrame depende, portanto, do centro médico em que o paciente é tratado. Para comparar centros pelo procedimento de Marascuilo, calcule, primeiramente, as proporções de derrame recorrente em cada centro médico.

Derrame recorrente	**Centro médico**				**Total**
	A	**B**	**C**	**D**	
Sim	16	12	21	12	61
Não	179	70	78	54	381
Total	195	82	99	66	442
Proporção com recorrência	0,082	0,146	0,212	0,182	0,138

Aplicando a fórmula (6.2), calcule as amplitudes críticas. Primeiro, é preciso obter o valor crítico de $\chi2$ com 3 graus de liberdade, no nível de significância estabelecido (Tabela 2 do Apêndice). Para 5%, esse valor é 7,82. Calcule, então, a raiz quadrada desse valor, que é é 2,796. Para obter as amplitudes críticas, faça os diversos cálculos intermediários. Ache, depois, as diferenças entre as proporções, duas a duas, e compare com as respectivas amplitudes críticas, para estabelecer as significâncias.

Cálculos intermediários para obtenção do valor da amplitude crítica para o procedimento de Marascuilo

Cálculo	**A**	**B**	**C**	**D**
n	195	82	99	66
p	0,082	0,146	0,212	0,182
1 – p	0,918	0,854	0,788	0,818
p (1 – p)	0,075319	0,124926	0,167126	0,14876
p (1 – p) / n	0,000386	0,001523	0,001688	0,002254

Tabela resumo para o procedimento de Marascuilo

Comparação	**Diferença entre proporções**	**Amplitude crítica**	**Significância**
A e B	0,06429	0,122205	Não
A e C	0,13007	0,127365	Sim

A e D	0,099767	0,143688	Não
B e C	0,06578	0,158477	Não
B e D	0,035477	0,171871	Não
C e D	0,030303	0,175576	Não

A proporção de derrames recorrentes no Centro C foi significativamente maior do que no Centro A.

6.4. TESTE DE χ2 DE MANTEL-HAENSZEL

O teste de χ2 de Pearson é aplicado para estabelecer se existe associação estatística entre duas variáveis. É possível, porém, que um fator estranho ao problema interfira sobre uma das variáveis. Lembre-se dos estudos sobre associação do hábito de fumar e câncer do pulmão, como aquele relatado no Exemplo 5.3 do Capítulo 5. Se não tivessem sido tomados todos os cuidados metodológicos, a diferença estatística entre fumantes e não fumantes poderia ter sido mascarada por outros fatores, como idade, por exemplo. Isso teria acontecido se pessoas mais velhas, que fumam menos e, ao mesmo tempo, têm mais risco de doenças, tivessem formado o grupo de não fumantes. Neste caso, diríamos que idade foi um *fator interveniente*, porque teria mascarado o efeito do tabagismo sobre o câncer do pulmão.

Quando existe um fator interveniente, a solução é "ajustar" o efeito desse fator, fazendo uma estratificação. No caso de a idade estar interferindo no hábito de fumar, a solução seria formar diferentes grupos de idade e, em cada grupo, contar quantos fumam e quantos não fumam. Em outras palavras, a solução seria organizar diversas tabelas 2 $\frac{3}{2}$ 2, uma para cada grupo de idade e, depois, combinar os resultados em uma só tabela. É isso que faz o teste de χ2 de Mantel-Haenszel.

Indicação: O teste de χ2 de Mantel-Haenszel é um procedimento para estudar a associação de duas variáveis quando é necessário fazer o ajuste para um fator interveniente.

Exemplo 6.7

Um pesquisador acompanha um grande grupo de infartados há 5 anos e tem os dados sobre atividade física e a eventual recorrência do infarto nesse grupo. Para saber se a recorrência de infarto está associada à atividade física regular, organizou os dados em uma tabela 2 × 2.

Participantes da pesquisa segundo o hábito de atividade física regular e infarto recorrente

Atividade física regular	**Infarto recorrente**	
	Sim	**Não**
Sim	60	380
Não	90	350

No entanto, o pesquisador suspeita de que a gravidade do primeiro infarto seria um fator interveniente, ou seja, suspeita de que a ocorrência de um segundo infarto possa depender da gravidade do primeiro. Se essa hipótese for verdadeira, a validade do teste de χ2 de Pearson fica prejudicada porque os grupos em comparação podem ser constituídos de pessoas que tiveram infartos com gravidades muito diferentes.

O pesquisador separa, então, os casos em três grupos, pela gravidade do primeiro infarto: leve, moderado, severo. Organiza, assim, três tabelas 2 × 2: a primeira apresenta casos de primeiro infarto leve, a segunda, de primeiro infarto moderado e a terceira, de primeiro infarto severo. Depois, agrupa as três tabelas em uma só.

Participantes da pesquisa segundo a gravidade do primeiro infarto, o hábito de atividade física regular e infarto recorrente

Gravidade do infarto	**Atividade física regular**	**Infarto recorrente**		**Total**
		Sim	**Não**	
Leve	Sim	10	70	80
	Não	10	60	70
Moderada	Sim	20	160	180
	Não	40	130	170
Severa	Sim	30	150	180
	Não	40	160	200
Total		150	730	880

Para aplicar o teste de Mantel-Haenszel:[85]

Primeiro passo: Estabeleça as hipóteses e o nível de significância. Então:

- Hipótese da nulidade: as variáveis são independentes.
- Hipótese alternativa: as variáveis *não* são independentes.
- Nível de significância: α = 5%.

Segundo passo: Organize não uma, mas tantas tabelas 2 × 2 quantos forem os níveis do fator interveniente. Por exemplo, se você estiver estudando a proporção de pessoas com câncer de pulmão em fumantes e não fumantes, mas suspeita de que a idade possa ser um fator interveniente, divida os participantes de pesquisa em *k* grupos de idade, antes de separar fumantes de não fumantes, com a doença e sem a doença. Você terá, em cada grupo de idade, os dados a_i; b_i; c_i; d_i; $i = 1, 2, ..., k$. Veja a notação literal e simplificada para tabelas 2 × 2, dada na Tabela 5.1 do Capítulo 5. Vamos repeti-la aqui, como Tabela 6.1, indicando os *k* estratos (no exemplo, grupos de idade).

Tabela 6.1 – Valores literais em tabela 2 × 2 estratificada

Fator interveniente	Grupo	Característica	
		Não	Sim
1	A	a_1	b_1
	B	c_1	d_1
2	A	a_2	b_2
	B	c_2	d_2
. . .			
k	A	a_k	b_k
	B	c_k	d_k

A fórmula de cálculo para o valor não corrigido de χ2 de Pearson pode ser escrita da seguinte

forma:

$$\chi^2 = \frac{(ad - bc)^2 n}{(a+b)(c+d)(a+c)(b+d)} \tag{6.5}$$

Terceiro passo: Para aplicar o teste de χ2 de Mantel-Haenszel, a fórmula é:

$$\chi^2_{MH} = \frac{\left\{\sum\left[a_i - \frac{(a_i+b_i)(a_i+c_i)}{n_i}\right]\right\}^2}{\sum \frac{(a_i+b_i)(c_i+d_i)(a_i+c_i)(b_i+d_i)}{n_i^2(n_i-1)}} \tag{6.6}$$

Quarto passo: Sob a hipótese de independência, ou seja, de que não existe associação entre as variáveis desde que corrigido o fator interveniente, a estatística calculada tem distribuição aproximada de χ2 com 1 grau de liberdade. Compare o valor calculado de χ2 com o valor crítico com 1 grau de liberdade e no nível estabelecido de significância (Tabela 2 do Apêndice).

Exemplo 6.8

Vamos aplicar o teste de χ2 de Mantel-Haenzel para os dados do exemplo 6.7. Reveja a fórmula. Faça:

$$e_i = \frac{(a_i+b_i)(a_i+c_i)}{n_i}$$

$$h_i = \frac{(a_i+b_i)(c_i+d_i)(a_i+c_i)(b_i+d_i)}{n_i^2(n_i-1)}$$

A estatística de teste é:

$$\chi^2_{MH} = \frac{\left[\sum(a_i - e_i)\right]^2}{\sum h_i}$$

Cálculos intermediários para obtenção do χ2 de Mantel-Haenszel

	Gravidade do primeiro infarto			Total
	Leve	Moderada	Severa	
ai	10	20	30	
ei	10,666667	30,857143	33,157895	
(ai – ei)	–0,666667	–10,857143	–3,157895	–14,6817
hi	4,343028	12,454009	14,274333	31,0714

$$\chi^2_{MH} = \frac{(-14,6817)^2}{31,0714} = 6,937$$

Como o valor calculado de χ2 é maior do que o valor crítico com 1 grau de liberdade e no nível de significância de 5% (Tabela 2 do Apêndice), conclui-se que existe associação estatística entre atividade física regular e recorrência de infarto.

Para determinar a direção da associação, foram calculados os percentuais de casos de infarto recorrente entre pessoas que têm e que não têm o hábito de realizar atividade física regularmente. É fácil ver, observando os percentuais de infarto recorrente, que atividade física faz diminuir significantemente o risco de infarto recorrente.

Percentuais de casos de infarto recorrente

Atividade física regular	**Gravidade do primeiro infarto**		
	Leve	**Moderada**	**Severa**
Sim	12,5	11,1	16,7
Não	14,3	23,5	20

6.5. TESTE DE χ2 PARA TENDÊNCIA

Indicação: o teste de χ2 de Pearson é indicado para avaliar a *tendência* quando os participantes da pesquisa são classificados segundo duas variáveis, mas uma delas é ordinal ou numérica. Se a ordenação dos dados não for levada em conta,[86] a análise estará errada.

Exemplo 6.9

A cárie dentária é uma doença que afeta a todos, indiscriminadamente. A prevenção é extremamente eficaz. No entanto, nas periferias, a incidência de cárie é alta porque falta prevenção. Para mostrar que na periferia de Brasília a proporção de crianças com dentes cariados é alta, uma pesquisadora levantou uma amostra de crianças e as classificou segundo a idade em anos completos e o fato de ter ou não cárie. Evidentemente, o número de crianças com dentes cariados aumenta com a idade. Mesmo assim, vamos usar esses dados para mostrar como se faz o teste de χ2 se houver tendência. É dada a proporção de crianças sem cárie em função da idade em tabela e em diagrama de dispersão. A primeira variável é discreta e a segunda é nominal.

Crianças com e sem experiência de cárie segundo a idade em anos completos

Cárie	**Idade em anos completos**					
	7	**8**	**9**	**10**	**11**	**12**
Sim	34	56	72	82	92	280
Não	10	14	16	14	12	20
Total	44	70	88	96	104	300
Proporção	0,227	0,200	0,182	0,146	0,115	0,067

Fonte: RIBEIRO, L.D.S.V. Estudo epidemiológico da prevalência de cárie dentária em crianças da cidade satélite de Guará – DF-Brasília. Dissertação. C.P.O. São Leopoldo Mandic, 2003.

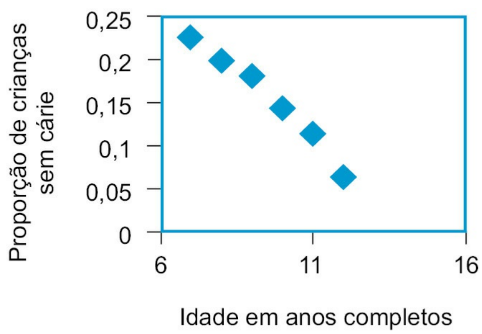


Figura 6.1 – Proporção de crianças sem cárie, segundo a idade, em anos.

Como o número de crianças sem cáries diminui à medida que a idade aumenta, é possível ajustar aos dados uma reta, ou seja, fazer uma *análise de regressão*. No entanto, o procedimento *não* seria simples porque é necessário levar em conta o fato de as proporções terem sido estimadas com base em números diferentes de crianças.

De qualquer forma, existe uma alternativa, que mostraremos aqui usando o Exemplo 6.8: fazer um teste de χ2 para tendência. Para isso:

Primeiro passo:

• Hipótese da nulidade: a proporção de crianças sem cáries não depende da idade.
• Hipótese alternativa: a proporção de crianças sem cáries diminui em função da idade.
• Nível de significância: 5%.

Segundo passo: Calcule o valor de χ2 usando a fórmula (6.1). O valor calculado de χ2 para os dados do Exemplo 6.8, com 5 graus de liberdade, é 20,52. O valor crítico de χ2 com 5 graus de liberdade e no nível de significância de 5% é 11,07 (Tabela 2 do Apêndice). Logo, a proporção de crianças sem cárie depende da idade.

Terceiro passo: Identifique a *variável ordinal ou numérica* do problema. No Exemplo 6.8, a variável idade é numérica, discreta. Atribua um posto (número de ordem) a cada categoria dessa variável, começando com 1 e aumentando, de unidade em unidade, até *s* (*s* é o número de categorias dessa variável). Indique os escores por *xi*, em que *i* = 1, 2, 3,..., *s*.

Quarto passo: Indique o *número total de indivíduos* em cada categoria da nova variável por n_{ii}, em que *i* = 1, 2, 3,...*s*. Evidentemente, õ*ni* = *n*, que é o tamanho da amostra. No Exemplo 6.8:

n = õn_i = 44 + 70 + 88 + 96 + 104 + 300 = 702

Indique o *número de indivíduos com a característica de interesse* em cada uma das categorias da variável ordinal por r_i, em que *i* = 1, 2, 3,... *s*. No Exemplo 6.8, r_i indica o número de crianças sem cárie em cada idade.

Logo:

õ*ri* = 10 + 14 + 16 + 14 + 12 + 20 = 86

Cálculos intermediários para aplicar o teste de χ2 para tendência

Estatísticas	**Idade**						**Total**
	7	**8**	**9**	**10**	**11**	**12**	
Escore	1	2	3	4	5	6	
Sem cárie	10	14	16	14	12	20	86
Total (n_i)	44	70	88	96	104	300	702

Quinto passo: Calcule a proporção de indivíduos com a característica de interesse em toda a amostra. Indique por *p*. Então:

$$p = \frac{\Sigma r_i}{n} \qquad (6.3)$$

No Exemplo 6.8:

$$p = \frac{\sum r_i}{n} = \frac{86}{702} = 0{,}1225$$

Sexto passo: Calcule a estatística:[87]

$$\chi^2_{tend} = \left(\frac{\sum r_i x_i - p \sum n_i x_i}{p(1-p)\left[\sum n_i x_i - \frac{\left(\sum n_i x_i\right)^2}{n} \right]} \right)^2 \tag{6.4}$$

O valor calculado de $\chi 2_{\ tend}$ tem, aproximadamente, distribuição de $\chi 2$ com 1 grau de liberdade.

Continuação dos cálculos intermediários para aplicar o teste de $\chi 2$ para tendência

Estatísticas	Idade						Total
	7	8	9	10	11	12	
Escore (x_i)	1	2	3	4	5	6	
Sem cárie (r_i)	10	14	16	14	12	20	ōr_i = 86
Total (n_i)	44	70	88	96	104	300	n = 702
$r_i x_i$	10	28	48	56	60	120	ō$r_i x_i$ = 322
$n_i x_i$	44	140	264	384	520	1.800	ō$n_i x_i$ = 3152
$n_i x_i^2$	44	280	792	1.536	2.600	10.800	ō$n_i x_i^2$ = 16.052

Agora, é possível calcular:

$$\chi^2_{tend} = \frac{322 - 0{,}1225 \times 3152^2}{0{,}1225\,(1 - 0{,}1225)\left[16052 - \frac{3152^2}{702}\right]}$$

$$\chi^2_{tend} = \frac{(-64{,}15152)^2}{0{,}10750 \times 1899{,}430} = 20{,}1550$$

Como o valor calculado de $\chi 2_{tend}$ com 1 grau de liberdade é maior do que o valor crítico no nível de 5% de significância com 1 grau de liberdade (Tabela 2 do Apêndice), a conclusão é a de que a proporção de crianças sem cáries diminui linearmente com a idade.

Sétimo passo: Calcule a diferença entre o $\chi 2$ de Pearson obtido pela fórmula (6.1) e o valor de χ^2_{tend} obtido pela fórmula (6.4).

Para o Exemplo 6.8, o teste de $\chi 2$ de Pearson calculado anteriormente resultou em 20,52, com 5 graus de liberdade. O teste de χ^2_{tend} resultou em 20,15, com 1 grau de liberdade. A diferença entre esses dois valores está associada a 5 – 1 = 4 graus de liberdade e é:

$$20,52 - 20,15 = 0,37$$

Essa diferença tem, aproximadamente, distribuição de $\chi 2$. Serve para testar a hipótese de que não existe outra causa de variação, além da explicada pela tendência, para a variação da variável ordenada. Como o valor calculado é não significante no nível de 5%, conclui-se que a reta é apropriada para descrever a variação da proporção de crianças sem cáries em função da idade.

IMPORTANTE:

O valor do χ^2_{tend} é sempre menor do que o do $\chi 2$ convencional. O primeiro teste está associado a 1 grau de liberdade e o segundo está associado a (s – 1) graus de liberdade. O primeiro é *mais poderoso*, ou seja, tem maior probabilidade de rejeitar a hipótese da nulidade quando essa hipótese é falsa.

RESUMO E OBJETIVO DO CAPÍTULO

Depois de estudar este Capítulo, você deve ser capaz de aplicar e interpretar:

1. Teste de $\chi 2$ para tabelas de contingência $2 \times s$, bem como partição dessas tabelas.
2. Procedimento de Marascuilo.
3. Teste de $\chi 2$ de Mantel-Haenszel para tabelas 2×2 agregadas.
4. Teste de $\chi 2$ para tendência.

---

[83] O teste não mostrou diferença entre as drogas, o que não significa que as drogas tenham o

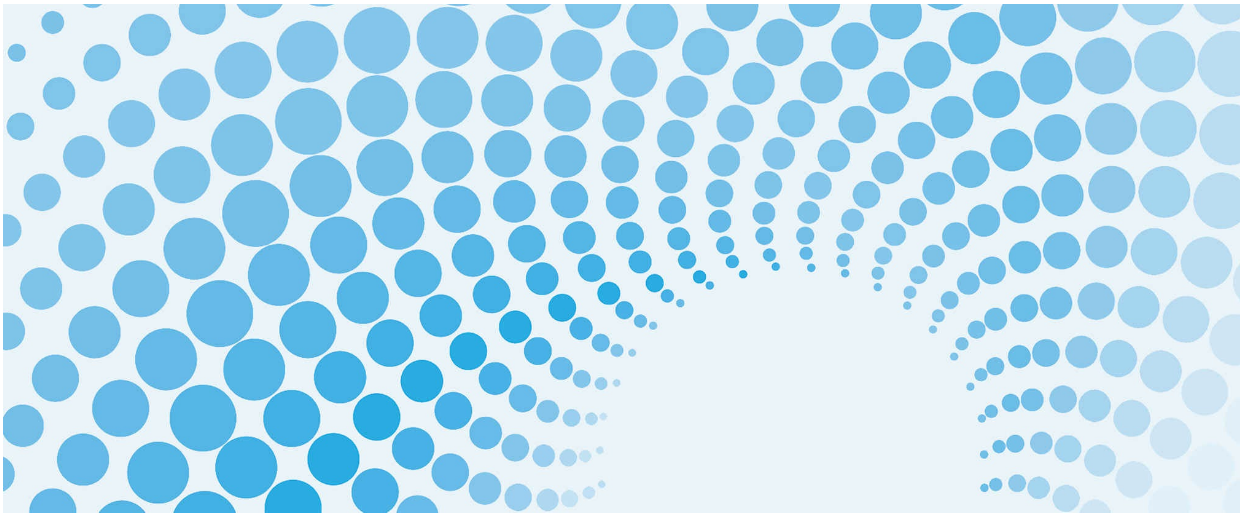
mesmo efeito.

[84] MARASCUILO, L.A.; SERLIN, R.C. *Statistical Methods for the Social and Behavioral Sciences*. Nova York. Freeman, 1988.

[85] Veja SCHORK, M.A.; REMINGTON, R.D. *Statistics with Applications to the Biological and Health Sciences.* 3. ed. Upper Saddle River: Prentice Hall, 2000. p. 216-220.

[86] Veja ALTMAN, D.O. *Practical Statistics for Medical Research.* 2nd ed. London: Chapman & Hall, 1993. p. 261-265.

[87] Para demonstração da estatística a partir de uma regressão, veja FLEISS, J. *Statistical methods for rates and proportions.* 2nd ed. New York: Wiley, 1981. p. 144.



6.6.1. Para verificar se a distribuição dos tipos sanguíneos independe da origem, foram obtidos os dados apresentados na Tabela 6.2. Faça o teste de χ2.

Tabela 6.2 – Participantes da pesquisa segundo a origem e o tipo sanguíneo

Origem	**Tipo sanguíneo**			
	O	**A**	**B**	**AB**
Árabe	130	149	29	8
Não árabe	417	292	94	17

Fonte: GARCIA, M.V. Blood groups and dermatogliphics in a Brazilian population of arabian origin. *Cienc Cult* 29(7):826-829, 1977.

6.6.2. Foi feito um estudo transversal de 976 formulários de registro de dados dos casos atendidos pelo SOS Criança, concluídos em 1993. Desses, 587 referiam-se a denúncias de maus-tratos. As denúncias foram classificadas em confirmadas e não confirmadas, segundo o perfil do notificante. Os dados estão na Tabela 6.3. Faça o teste de χ2.

Tabela 6.3 – Denúncias confirmadas e não confirmadas segundo o perfil do notificante

Perfil do notificante	**Denúncia**		
	Não confirmada	**Confirmada**	**Total**
Familiares	141	50	191
Amigos e vizinhos	140	22	162
Anônimo	84	14	98
Desconhecido	55	6	61
Profissionais	22	3	25
A própria criança	20	2	22
Outros	24	4	28
Total	486	101	587

Fonte: GONÇALVES, H.S. et al. Avaliação de serviço de atenção a crianças vítimas de violência doméstica. *Rev Saúde Pública* 33(6): 547-553, 1999.

6.6.3. Para estudar se existe associação entre disfunção craniomandibular e nível de renda em portadores de próteses duplas, um pesquisador levantou os dados apresentados na Tabela 6.4. Algumas frequências esperadas podem ter, no entanto, valor menor do que 5. Verifique e, se isso realmente ocorrer, dê uma sugestão de procedimento ao pesquisador.

Tabela 6.4 – Distribuição de portadores de próteses duplas segundo o nível de renda, em salários mínimos (SM), e o grau de disfunção craniomandibular

Nível de renda	**Grau da disfunção craniomandibular**			
	Nenhuma	**Leve**	**Moderada**	**Severa**
Menos de 5 SM	19	21	10	0
De 5 a 10 SM	21	24	5	0
Mais de 10 SM	25	21	2	2

Fonte: MOLLO JR., F.A. et al. Influência do nível socioeconômico sobre a prevalência de disfunção craniomandibular em pacientes portadores de dentaduras duplas. *Rev Bras Prot Clin Labor* 2(8): 62-65 [s.d].

6.6.4. Escolha a alternativa correta para completar a frase: O teste de χ2 é usado para

a) estudar a linearidade das variáveis

b) obter a variância

c) verificar se as variáveis são binomiais

d) testar a hipótese de independência

e) calcular a correlação

6.6.5. Para comparar a eficácia e a tolerabilidade da combinação das drogas candesartan cilexetil e hidroclorotiazida com a combinação das drogas losartan e hidroclorotiazida na hipertensão, foi feito um ensaio clínico com 271 pacientes e foram obtidos os dados apresentados na Tabela 6.5. Por "paciente controlado" entende-se aquele que, na última consulta, tinha PAD (pressão arterial diastólica) menor que 90 mmHg na posição sentado e por "respondedor" aquele que tinha PAD maior que 90 mmHg, mas havia conseguido diminuição maior do que 10 mmHg entre a primeira e a última consulta. Por paciente "não respondedor" entende-se aquele que não conseguiu PAD menor que 90 mmHg, nem diminuição da PAD maior do que 10 mmHg entre a primeira e a última consulta. Faça o teste de $\chi 2$. Depois, faça a partição.

Tabela 6.5 – Pacientes segundo a resposta à combinação de hidroclorotiazida com uma segunda droga

Segunda Droga	**Resposta do paciente**			**Total**
	Controlados	**Respondedores**	**Não Respondedor**	
Candesartan	85	10	44	139
Losartan	65	10	57	132
Total	150	20	101	271

Fonte: OHMAN, K.P.; MILON, H.; VALNES, K. Eficácia e tolerabilidade do comprimido da combinação de Candesartan Cilexetil e Hidroclorotiazida na hipertensão essencial insuficientemente controlada – comparação com a combinação de Losartan e Hidroclorotiazida. *Blood Pressure* 9:214-220, 2000.

6.6.6. A ANVISA (Agência Nacional de Vigilância Sanitária) disponibiliza na internet um questionário básico sobre a percepção de profissionais da saúde a respeito das infecções relacionadas à assistência à saúde e à higienização das mãos. A primeira questão não demográfica que trata do assunto em questão é: "Você recebeu algum treinamento em higienização das mãos?". Imagine que esse questionário foi entregue a 1.500 profissionais de diferentes tipos, que trabalham em diversas instituições de uma grande metrópole. As respostas, segundo o tipo de profissional, estão na Tabela 6.6. Faça o teste de $\chi 2$. Depois, considerando o tipo do profissional, compare a proporção de respostas "Não" usando o procedimento de Marascuilo.

Tabela 6.6 – Respostas segundo o tipo de profissional

Tipo de profissional	**Resposta**	
	Sim	**Não**
Enfermeiro (E)	264	36
Auxiliar de enfermagem (A)	254	46
Parteira (P)	258	42
Técnico (T)	237	63
Outros (O)	262	38

6.6.7. A Tabela 6.7 apresenta os dados de uma amostra de 178 usuários de lentes de contato classificados de acordo com o sexo e a idade em que começaram a usar esse tipo de lente. Os dados trazem alguma evidência de que sexo e idade em que se começa a usar lentes de contato são dependentes?

Tabela 6.7 – Usuários de lentes de contato classificados de acordo com o sexo e a idade em que começaram a usar esse tipo de lente

Faixa de idade	Sexo		Total
	Homens	Mulheres	
Menos de 15 anos	2	8	10
De 15 a 19 anos	38	93	131
20 anos e mais	22	15	37
Total	62	116	178

Fonte: BAILEY, N.J. 1968. Contact Lenses design – a survey. Amer J Optom and Arch Amer Acad Optom 45: 96. Apud DANIEL, W.W. Applied Nonparametric Statistics. Pacific Grove: Duxbury, 2000.

6.6.8. Para avaliar o conhecimento de estudantes de Medicina e de médicos sobre Ética Médica relacionada à AIDS, foi aplicado um questionário a 120 estudantes do curso médico e a 50 médicos. Os estudantes foram divididos em dois grupos, 60 do curso básico e 60 do curso clínico, e os médicos também em dois grupos, 20 residentes e 30 pós-residentes. À questão "O médico cirurgião HIV-positivo pode continuar exercendo sua profissão?" foram obtidas as respostas[88] apresentadas na Tabela 6.8. Faça o teste de $\chi 2$.

Nota: esta é uma tabela 4 × 3; você aplica a fórmula (6.1) para encontrar o valor de $\chi 2$.

Tabela 6.8 – Respostas à questão: "O médico cirurgião HIV-positivo pode continuar exercendo sua profissão?"

Resposta	Respondente				Total
	Estudante		Médico		
	Básico	Clínico	Residente	Pós-residente	
Sim	38	49	16	23	126
Não	17	8	2	6	33
Não declarou	5	3	2	1	11
Total	60	60	20	30	170

Fonte: SANTOS, O.S.; MENDES, E.V.; UMESAKI, M.C.; SANTOS, C.S.; AMBRÓSIO, M.R. Conhecimento de ética médica relacionada à AIDS entre estudantes de Medicina e médicos da Faculdade de Medicina da Universidade Federal de Uberaba. *Revista Bioética 17*(1):123-134, 2009.

6.6.9. São apresentados na Tabela 6.9 dados relativos ao tamanho do carcinoma de células renais e o tipo de detecção, se incidental ou por sintomas. a) Existe associação entre tamanho do tumor e tipo de detecção? b) À medida que aumenta o tamanho do tumor, aumenta a tendência de detecção?

Tabela 6.9 – Carcinoma de células renais segundo o tamanho do tumor, em centímetros, e o tipo de detecção

Tamanho	Tipo de detecção	
do tumor	Incidental	Sintomático
De 0,5 a 4	30	11
De 4 a 7	21	22
De 7 a 10	5	14
Maior que 10	3	9
Total	59	56

Fonte: DALLI'OGLIO, M.F. et al. Características morfológicas dos carcinomas de células renais incidentais. *Urologia Clínica.* São Paulo: Escola

Paulista de Medicina [s.d.].

6.6.10. Para verificar a associação entre mononucleose infecciosa e amidalectomia, foram levantados os dados apresentados na Tabela 6.10. O teste deve ser ajustado para a idade, uma vez que os participantes que não tinham a doença eram mais velhos. Faça o teste.

Tabela 6.10 – Participantes da pesquisa segundo idade, doença e amidalectomia

Idade	**Mononucleose infecciosa**	**Amidalectomia**	
		Sim	**Não**
	Sim	6	17
18	Não	17	32
	Sim	3	39
19	Não	26	70
	Sim	12	29
20	Não	34	78
	Sim	8	38
21	Não	48	91
	Sim	5	10
22	Não	45	73
	Sim	2	7
23	Não	29	37
	Sim	4	5
24	Não	36	39

Fonte: MILLER, R.G. ; EFRON, B. ; BROWN, B.W. ; MOSES, L.E. *Biostatistics casebook*. Wiley, 1980

[88] De acordo com os autores, o médico pode continuar exercendo sua profissão, mas precisa adotar normas de proteção e segurança.

7

Medidas de Associação e Correlação de Spearman

Medidas de associação são coeficientes que medem a *força da relação estatística* entre variáveis, mas *não* medem a significância estatística. Algumas associações são conhecidas e já foram absorvidas. Por exemplo, as pessoas associam, sem maiores discussões, tabagismo ao câncer de pulmão e risco de acidente de carro ao motorista alcoolizado. Outras vezes, a associação entre duas variáveis gera polêmica. Lembre-se da questão dos resultados no Exame Nacional do Ensino Médio (ENEM) e do desempenho dos estudantes na universidade. Existe interesse em estudar a associação entre essas variáveis, mas não existe consenso na interpretação dos resultados.

7.1. MEDIDAS DE ASSOCIAÇÃO EM TABELAS 2 × 2

Para medir o grau de associação entre duas variáveis nominais apresentadas em tabelas 2 × 2, vamos ver duas medidas de associação: o coeficiente fi e o coeficiente gama.

7.1.1. Coeficiente FI

O coeficiente fi, que se indica pela letra grega φ (lê-se fi), é muito usado por pesquisadores das áreas de psicologia e sociologia. Funciona bastante bem quando o estudo é transversal, mas não é tão bom no caso de estudos prospectivos e retrospectivos. Esse coeficiente é definido pela fórmula:

$$\varphi = \sqrt{\frac{\chi^2}{n}} \qquad (7.1)$$

em que χ^2 é o valor obtido pelo teste não corrigido de χ^2 e *n* é o tamanho da amostra. O valor do coeficiente φ varia entre 0 e 1, isto é,

$$0 \le \varphi \le 1.$$

Você interpreta o resultado do coeficiente φ da seguinte forma:

1. Quanto mais próximo de 1 estiver o valor de φ, maior será o grau de associação entre as variáveis.
2. Quanto mais próximo de 0 (zero) estiver o valor de φ, menor será a associação entre as variáveis.
3. $\varphi = 1$ significa *associação perfeita*.[89]
4. $\varphi = 0$ significa *associação nula*.
5. Como regra prática, valores de φ menores do que 0,30 ou 0,35 podem ser tomados como indicadores de associação trivial[90] entre as duas variáveis.

Você deve estar se perguntando: devo calcular um coeficiente de associação, se já apliquei um teste de χ^2 para verificar a significância dos resultados? A resposta é sim. Existe a possibilidade de o teste estatístico ser significante e a associação entre as variáveis ser fraca. Isso acontece porque o teste de χ^2 – como todo teste, aliás – sofre a influência do *tamanho da amostra*. Quanto maior for a amostra, maior será a probabilidade de encontrar resultados significantes.

Por outro lado, o grau de associação é função *somente das proporções observadas*. Se a amostra for pequena, existe a possibilidade de o teste estatistico ser não significante e haver associação entre variáveis. Pense em uma tabela de contingência 2 × 2: se você mantiver as proporções e aumentar o tamanho da amostra, não mudará o o grau de associação, mas estará

aumentando a probabilidade de significância estatística.

Exemplo 7.1

Foi feito um levantamento[91] para verificar se a anodontia parcial (ausência congênita de um ou mais dentes) está associada ao sexo. Como os dados foram obtidos por estudo transversal, é recomendável calcular a distribuição percentual.

Escolares segundo sexo e anodontia parcial (ausência congênita de um ou mais dentes) (porcentagens entre parênteses)

Sexo	**Anodontia parcial**		**Total**
	Sim	**Não**	
Masculino	5 (1%)	270 (54%)	275 (55%)
Feminino	10 (2%)	215 (43%)	225 (45%)
Total	15 (3%)	485 (97%)	500 (100%)

Para testar a hipótese de que anodontia parcial independe de sexo, aplica-se o teste de χ^2. Usando a fórmula (5.1) do Capítulo 5, temos:

$$\chi^2 = \frac{(5\times 215 - 10\times 270)^2\, 500}{(5+270)(10+215)(5+10)(270+215)} = 2{,}93$$

O valor calculado de χ^2 com 1 grau de liberdade é menor do que o valor crítico no nível de 5% de significância (Tabela 2 do Apêndice). Não há, portanto, evidência de que anodontia parcial esteja associada ao sexo. Um programa para computador fornece *p*-valor igual a 0,0868.

O coeficiente de associação φ, calculado pela fórmula (7.1), é

$$\varphi = \sqrt{\frac{2{,}93}{500}} = \sqrt{0{,}00587} = 0{,}0766$$

O grau de associação medido pelo coeficiente φ entre sexo e anodontia parcial é pequeno, trivial. Então, com base nessa amostra, não há razão para considerar que o risco de anodontia parcial esteja associado ao sexo.

Para fazer os cálculos, você sempre pode usar um programa de computador. Note, porém, que alguns programas calculam φ^2 (*phi-square*), ou seja, o valor de fi ao quadrado – e não o valor de φ apresentado aqui. Consulte sempre o manual do programa para saber que estatísticas estão sendo calculadas. Usando o programa *Statistica*, você obterá, para os dados do Exemplo 7.1:

	Column 1	**Column 2**	**Row Totals**
Frequencies, row 1	5	270	275
Percent of total	1,000%	54,000%	55,000%
Frequencies, row 2	10	215	225
Percent of total	2,000%	43,000%	45,000%
Column totals	15	54,000%	55,000%
Percent of total	3,000%	97,000%	
Chi-square (df=1)	2,93	p = ,0868	

V-square (df=1)	2,93	p = ,0871
Yates corrected Chi-square	2,1	p = ,1473
Phi-square	0,00587	

Exemplo 7.2

Reveja o Exemplo 7.1. Considere agora que o pesquisador tenha obtido uma amostra *não* de 500 escolares, mas de 2.000, com a mesma distribuição percentual.

Distribuição de escolares segundo sexo e anodontia parcial

Sexo	Anodontia parcial		Total
	Sim	**Não**	
Masculino	20 (1%)	1.080 (54%)	1.100 (55%)
Feminino	40 (2%)	860 (43%)	900 (45%)
Total	60 (3%)	1.940 (97%)	2.000 (100%)

Para testar a hipótese de que anodontia parcial independe de sexo, aplica-se o teste de χ^2. Usando a fórmula (5.1) do Capítulo 5, vem:

$$\chi^2 = \frac{(20\times 860 - 1.080\times 40)^2\, 2.000}{(20+1.080)(40+860)(20+40)(1.080+860)} = 11,73$$

Como o valor calculado de χ^2 com um grau de liberdade é maior do que o valor crítico no nível de 5% de significância (Tabela 2 do Apêndice), pode-se afirmar que a anodontia parcial está associada ao sexo. Se você usar um programa para computador, obterá *p*-valor igual a 0,0006. O valor do coeficiente de associação φ obtido pela fórmula (7.1) é:

$$\varphi = \sqrt{\frac{11,73}{2.000}} = \sqrt{0,00587} = 0,0766$$

Veja bem: a associação entre anodontia parcial e sexo é apenas trivial – apesar de o resultado do teste de χ^2 ter sido significante no nível de 5%.

Compare os Exemplos 7.1 e 7.2: as proporções (dentro das células) são iguais nos dois exemplos, mas a amostra do Exemplo 7.2 é quatro vezes maior. E o que aconteceu com as estatísticas? O aumento da amostra fez aparecer a significância estatística, embora o grau de associação tenha permanecido o mesmo e – no caso – bastante pequeno.

Isso acontece porque, quando o tamanho da amostra aumenta, aumenta a probabilidade de encontrar valor estatisticamente significante. No entanto – desde que a distribuição seja a mesma (as mesmas proporções nas células) –, o grau de associação entre as variáveis permanecerá o mesmo. Portanto, cuidado com a conclusão de estudos transversais com grandes amostras: a associação pode ser trivial, mesmo que haja significância estatística.

É importante observar que:

- O teste de $\chi 2$ testa *a hipótese* de que a associação entre duas variáveis nominais é *significante*.
- O coeficiente $\$$ mede o *grau de associação* entre duas variáveis nominais.
- Como são estatísticas diferentes – a primeira mede a *significância da associação* e a segunda, o *grau de associação* –, recomenda-se calcular as duas e, depois, discutir os resultados.

7.1.2. Coeficiente gama

O coeficiente gama,[92] que se indica pela letra grega $\circ$ (lê-se gama), mede o grau da associação com que *duas categorias ordenadas* de variáveis tendem a crescer – e, portanto, decrescer – juntas. É definido por:

$$\gamma = \frac{ad - bc}{ad + bc} \qquad (7.2)$$

em que *a*, *b*, *c* e *d* são os valores definidos na Tabela 5.1 do Capítulo 5. O valor do coeficiente gama fica entre –1 e +1, inclusive.

$$-1 \lfloor \circ \lfloor +1.$$

O coeficiente gama deve ser interpretado como segue:

- $\circ = 1$: *associação perfeita positiva.*
- $\circ = -1$: *associação perfeita negativa.*
- $\circ = 0$: *associação nula.*
- $0 < \circ < 1$: *associação positiva.*
- $-1 < \circ < 0$: *associação negativa.*

Exemplo 7.3

Foi feito um estudo transversal para verificar se a doença periodontal (gengivite) está associada ao hábito de fumar. Veja os dados:[93] as variáveis estão "ordenadas" em linha, de 0 (não tem a doença) a 1 (tem a doença); em coluna, de 0 (não fuma) a 1 (fuma).

Participantes de uma pesquisa classificados segundo o hábito de fumar e a doença periodontal (porcentagens entre parênteses)

Hábito de fumar	**Doença periodontal**		**Total**
	Não	**Sim**	
Não	18 (37,5%)	6 (12,5%)	24 (50,0%)
Sim	14 (29,2%)	10 (20,8%)	24 (50,0%)
Total	32 (66,7%)	16 (33,3%)	48 (100%)

Para testar a hipótese de que a doença periodontal[94] independe do hábito de fumar, aplica-se o teste de χ^2. Usando a fórmula (5.1) do Capítulo 5, temos:

$$\chi^2 = \frac{(18 \times 10 - 6 \times 14)^2\, 48}{(18+6)(14+10)(18+14)(6+10)} = 1{,}50$$

Como o valor calculado de χ^2 é menor do que o valor crítico com 1 grau de liberdade e no nível de 5% de significância (Tabela 2 do Apêndice), não se pode afirmar que a doença periodontal esteja associada ao hábito de fumar. Um programa para computador fornece *p*-valor igual a 0,2207.

Para obter o valor do coeficiente γ, aplica-se a fórmula (7.2). Deve-se lembrar que, neste exemplo, $a = 18$, $b = 6$, $c = 14$ e $d = 10$, temos:

$$\gamma = \frac{18 \times 10 - 6 \times 14}{18 \times 10 + 6 \times 14} = \frac{96}{264} = 0{,}36$$

Veja bem: embora o teste de χ^2 seja não significante no nível de 5%, o coeficiente γ aponta associação positiva, embora pequena, entre hábito de fumar e doença periodontal. É, porém, razoável considerar que uma amostra maior, neste caso, revelaria resultado significante. Afinal, o teste de χ^2 tem pouco poder[95] quando a amostra é pequena.

O coeficiente gama varia entre –1 e +1, inclusive, isto é, $-1 \leq \gamma \leq +1$. O coeficiente fi ϕ que vimos anteriormente, varia entre 0 e 1, isto é, $0 \leq \phi \leq 1$. Então, o coeficiente ϕ mostra o grau da associação entre duas variáveis nominais, mas o coeficiente γ fornece, também, o sentido da associação. Cuidado, portanto, ao desenhar a tabela para calcular o coeficiente γ, porque, invertendo as linhas, o sinal do coeficiente muda (e, evidentemente, a interpretação).

7.1.3. Razão de chances

A razão de chances (*OR*) é uma das várias estatísticas que se tornaram muito importantes na pesquisa clínica e na tomada de decisões. É particularmente útil porque fornece informação clara aos clínicos sobre qual é o tratamento que tem as melhores chances de beneficiar o paciente.

Por definição, *chance* é a razão entre a probabilidade (p) de determinado evento ocorrer e a probabilidade (q) de que esse evento não ocorra. Indica-se chance pela letra w.

$$w = \frac{p}{q} \tag{7.3}$$

Evidentemente, $p + q = 1$.

Exemplo 7.4

Lembre-se dos primeiros experimentos de Genética, conduzidos pelo monge Gregor Mendel. Ervilhas verdes cruzadas com ervilhas amarelas produziram ervilhas amarelas, que cruzadas entre si segregaram na proporção de três amarelas para cada verde. Então:

- “a probabilidade de ocorrer ervilha amarela quando se cruzam ervilhas amarelas heterozigotas é de ¾” e

• “a probabilidade de ocorrer ervilha verde é de ¼.”

A chance é

$$w = \frac{\frac{3}{4}}{\frac{1}{4}} = \frac{3}{1}$$

que se lê 3 para 1. Você também pode escrever 3:1.

Vamos ver, agora, um exemplo fictício com números inteiros para facilitar tanto o cálculo como a interpretação do resultado na área de saúde.

Exemplo 7.5

Imagine que um pesquisador que saber se existe associação entre a faixa de idade da mãe (categorizada como adolescente ou adulta) e o peso ao nascer do filho (categorizado como de baixo peso ao nascer ou de peso normal). Para isso, obteve, em uma grande maternidade, uma amostra casual simples dos prontuários de 100 mães adolescentes e 100 mães adultas, primíparas, com partos a termo de nascidos vivos.

Grupo de idade da mãe e peso ao nascer do filho

Grupo de idade da mãe	**Peso ao nascer**		**Total**
	Baixo	**Normal**	
Adolescente	40	60	100
Adulta	10	90	100
Total	50	150	200

A proporção de filhos de mães adolescentes com baixo peso ao nascer é:

$$p_1 = \frac{40}{100} = 0,4$$

A proporção de filhos de mães adolescentes com peso normal ao nascer é:

$$q_1 = \frac{60}{100} = 0,6$$

Estima-se, então, que a chance de baixo peso ao nascer para filhos de mães adolescentes seja:

$$w_1 = \frac{0,4}{0,6} = \frac{4}{6}$$

Lê-se “4 para 6”, ou seja, no caso de mães adolescentes a chance é de 4 bebês com baixo peso para 6 com peso normal. Veja a Figura 7.1.

A proporção de filhos de mães adultas com baixo peso ao nascer é:

$$p_2 = \frac{10}{100} = 0,1$$

A proporção de filhos de mães adultas com peso normal ao nascer é:

$$q_2 = \frac{90}{100} = 0,9$$

Estima-se, portanto, que a chance de a mãe adulta ter um bebê com baixo peso ao nascer é:

$$w_1 = \frac{0,1}{0,9} = \frac{1}{9}$$

Lê-se "1 para 9", ou seja, para mulheres adultas, a chance é 1 bebê com baixo peso para 9 com peso normal. Veja a Figura 7.1.

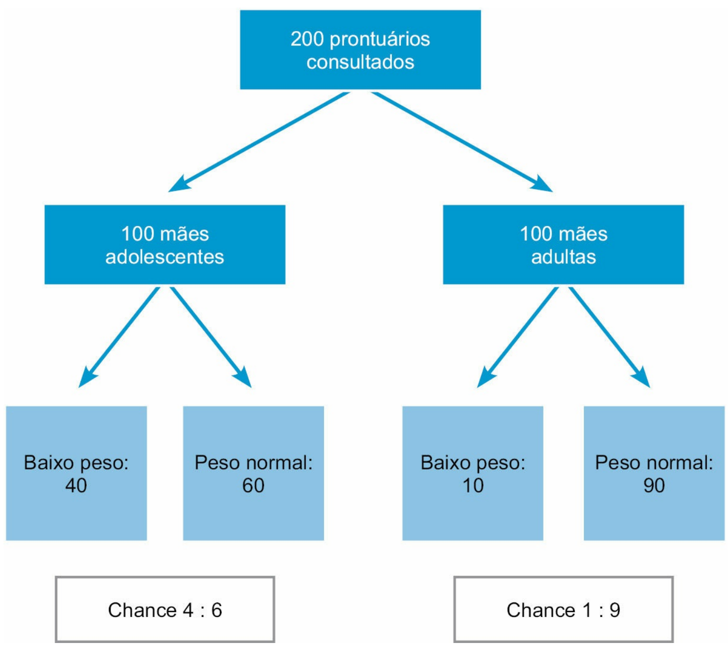


Figura 7.1 – Chances: peso ao nascer.

O interesse da pesquisa é comparar grupos. Essa comparação pode ser feita de diversas maneiras,[96] mas é comum calcular a *razão de chances.*

Razão de chances (*odds ratio*), que indicaremos pela letra *o* (de *odds*), é definida pela fórmula:

$$o = \frac{w_1}{w_2} \tag{7.4}$$

Exemplo 7.6

Reveja o Exemplo 7.5 e considere as chances calculadas. Aplicando a fórmula (7.4), a *razão de chances* é:

$$o = \frac{\frac{4}{6}}{\frac{1}{9}} = \frac{36}{6} = 6$$

O que significa razão de chances igual a 6? A chance de uma mãe adolescente ter filho com baixo peso ao nascer corresponde a 6 vezes a chance da mãe adulta.

Razão de chances

Grupo	Peso ao nascer		Total	Probabilidade		Chance	Razão de chance
	Baixo	Normal		Baixo	Normal		
Adolescente	40	60	100	0,4	0,6	0,666667	6
Adulta	10	90	100	0,1	0,9	0,111111	
Total	80	120	200				

A razão de chances é uma *medida de associação*. Saiba que:

1. Quando as duas variáveis são *independentes*, a razão de chances é igual a 1 (ou muito próximo de 1).
2. Quanto maior[97] for a razão de chances, maior será a associação entre as duas variáveis.
3. Não confunda *chance* com *probabilidade*. E, para que isso fique bem claro, lembre-se da primeira lei de Mendel: ervilhas amarelas heterozigotas cruzadas entre si segregaram na proporção de três amarelas para cada verde. Veja: você tem aí a *chance* de 3 para 1, mas a probabilidade de obter ervilha amarela quando se cruzam ervilhas amarelas heterozigotas é de ¾.

De qualquer modo, é difícil entender o que é, exatamente, razão de chances. E essa estatística tem aparecido muito em trabalhos da área médica. Vamos, então, ver outro exemplo, agora com dados reais.

Exemplo 7.7

Foi feito um estudo com 263 adolescentes que necessitaram de consulta psiquiátrica imediata após apresentarem comportamento suicida. Nos 6 meses seguintes, 186 desses adolescentes não mais apresentaram comportamento suicida. Desses 186 adolescentes, 86 haviam sido diagnosticados com depressão na primeira consulta psiquiátrica. Mantiveram comportamento suicida no período de acompanhamento de 6 meses 77 dos 263 adolescentes. Haviam sido diagnosticados com depressão na primeira consulta psiquiátrica 45 deles. Qual é a razão de chances para comportamento suicida, considerando o diagnóstico de depressão na primeira consulta psiquiátrica?

Dica: Nem sempre é fácil construir uma tabela olhando a descrição de um problema. Minha sugestão é rabiscar um gráfico. Veja a Figura 7.2.

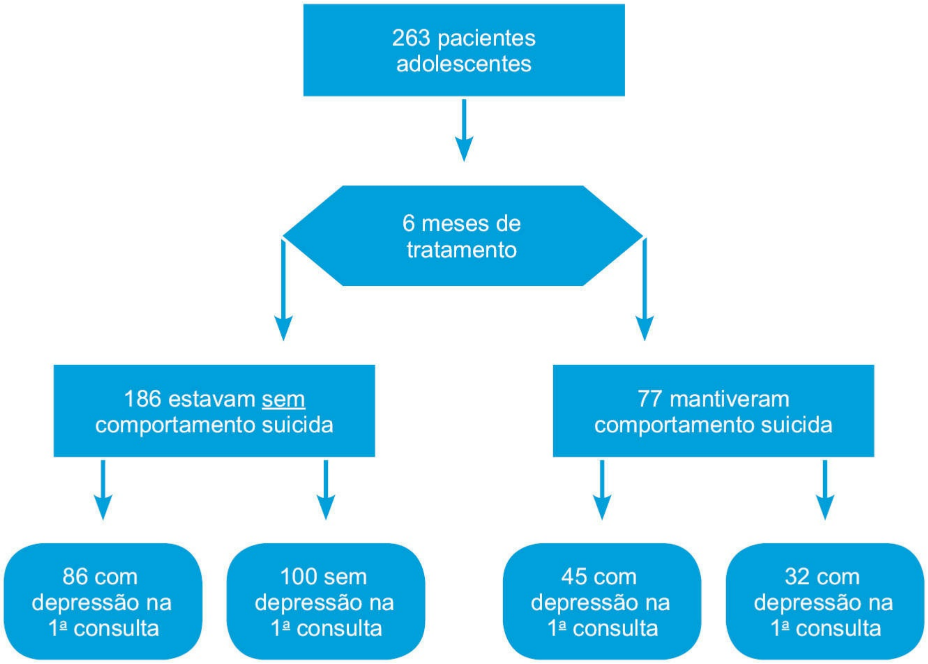


Figura 7.2 – Chances: sugestão de gráfico.

Distribuição dos adolescentes

Comportamento suicida	**Depressão**		**Total**
	Sim	**Não**	
Sim	45	32	77
Não	86	100	186
Total	131	132	263

Fonte: GREENFIELD, B.; HENRY, M., WEISS, M.; TSE, S.M.; GUILE, J.M.; DOUGHERTY, G.; ZHANG, X.; FOMBONNE, E.; LIS, E.; LAPALME-REMIS; HARNDEN, B. Previously suicidal adolescents: Predictors of six-month outcome. *Journal of the Canadian Association of Child and Adolescent Psychiatry* 17(4):197-201, 2008.

Vamos calcular as proporções de adolescentes com ou sem depressão na linha de base, que revelaram ou não comportamento suicida nos 6 meses de seguimento. As chances são calculadas usando a fórmula (7.3).

Chances de comportamento suicida persistente, em função de ter sido diagnosticada depressão na primeira consulta

Comportamento suicida	**Frequência**			**Proporção**			**Chance**
	Depressão			**Depressão**			
	Sim	**Não**	**Total**	**Sim**	**Não**	**Total**	
Sim	45	32	77	$p_1 = 0{,}584$	$q_1 = 0{,}416$	1	1,406
Não	86	100	186	$p_2 = 0{,}462$	$q_2 = 0{,}538$	1	0,859
Total	131	132	263				

A razão de chances é dada pela fórmula (7.4):

$$o = \frac{1,406}{0,859} = 1,63$$

A chance de comportamento suicida persistente é 1,63 maior se tiver sido diagnosticada depressão na primeira consulta psiquiátrica.

Exemplo 7.8

Foi levantada a prevalência de degeneração macular relacionada com a idade (DMRI) entre homens e mulheres com mais de 55 anos de idade que estavam sendo atendidos em dois hospitais de referência em Pernambuco. Os dados são apresentados em seguida.

Distribuição de pacientes atendidos em dois hospitais de Pernambuco segundo o sexo e o fato de serem portadores ou não de DMRI

Sexo	**DMRI**		**Total**
	Sim	**Não**	
Homens	46	94	140
Mulheres	61	199	260

Fonte: SANTOS, L.P.F. et al. DMRI: prevalência e fatores de risco em dois centros de referência em Pernambuco. *Arq Bras Oftalmol 68*(2): 229-233, 2005.

Calcule a proporção de homens com mais de 55 anos com degeneração macular relacionada com a idade (DMRI) e a proporção de homens com mais de 55 anos sem essa degeneração:

$$P(H, sim) = \frac{46}{140} = 0,3286$$

$$P(H, não) = \frac{94}{140} = 0,6714$$

A chance de homens com mais de 55 anos terem degeneração macular relacionada com a idade (DMRI) é, de acordo com a definição dada pela fórmula (7.3):

$$\omega_H = \frac{0,3286}{0,6714} = 0,4894$$

A chance de mulheres com mais de 55 anos de idade terem degeneração macular relacionada com a idade (DMRI) é, de acordo com a definição dada pela fórmula (7.3), 0,3065. Agora fica fácil calcular a razão de chances.

Razão de chances para DMRI em função do sexo

Sexo	**Frequência**			**Proporção**			**Chance**	**Razão de chances**
	DMRI			**DMRI**				
	Sim	**Não**	**Total**	**Sim**	**Não**	**Total**		
Homens	46	94	140	0,3286	0,6714	1	0,4894	1,5967 ≈ 1,6
Mulheres	61	199	260	0,2346	0,7654	1	0,3065	

Total	46	94	140

A razão de chances é de, aproximadamente, 1,6. O que isso significa? Para a proporção de 1,6 homem com mais de 55 anos que tem degeneração macular relacionada com a idade (DMRI) há 1 mulher na mesma condição, ou seja, há 16 homens para 10 mulheres na mesma condição (os cálculos foram aproximados).

A *razão de chances* é também conhecida como *razão dos produtos cruzados*. Há uma explicação para essa denominação. Veja: você calcula, para obter as probabilidades associadas a cada evento de uma tabela 2 × 2:

$$p_1 = \frac{a}{a+b}$$

$$q_1 = \frac{b}{a+b}$$

$$p_2 = \frac{c}{c+d}$$

$$q_2 = \frac{d}{c+d}$$

As chances para os dois grupos, 1 e 2, são dadas pela fórmula (7.3):

$$w_1 = \frac{\frac{a}{a+b}}{\frac{b}{a+b}}$$

$$w_2 = \frac{\frac{c}{c+d}}{\frac{d}{c+d}}$$

A razão de chances é dada pela fórmula (7.4). Então:

$$o = \frac{w_1}{w_2} = \frac{a}{a+b} \div \frac{b}{a+b} \div \frac{c}{c+d} \div \frac{d}{c+d} = \frac{a \times d}{b \times c}$$

Veja que a razão de chances fica definida pela fórmula

$$o = \frac{a \times d}{b \times c} \tag{7.5}$$

Então, para *obter a razão de chances*, use a fórmula (7.5), que é bem mais fácil.

Exemplo 7.9

Reveja os dados apresentados no Exemplo 7.8.

Distribuição de pacientes atendidos em dois hospitais de Pernambuco segundo o sexo e o fato de serem portadores ou não de DMRI

Sexo	DMRI		Total
	Sim	Não	
Homens	46	94	140
Mulheres	61	199	260

Fonte: SANTOS, L.P.F. et al. DMRI: prevalência e fatores de risco em dois centros de referência em Pernambuco. *Arq Bras Oftalmol* 68(2): 229-233, 2005.

Aplicando a fórmula (7.5), obtém-se a razão de chances:

$$o = \frac{46 \times 199}{61 \times 94} = \frac{9.154}{5.734} = 1{,}5964$$

7.1.4. Risco relativo

Nos estudos prospectivos, o pesquisador acompanha um grupo de pessoas com uma característica e um grupo de pessoas sem essa característica por determinado tempo, à espera da ocorrência de um desfecho. Depois, calcula a proporção de pessoas com o desfecho esperado nos dois grupos. Essas proporções são *estimativas de probabilidade*.

A probabilidade de perigo com ameaça física ou psicológica para o homem e/ou o meio ambiente é denominada *risco*. Nos estudos prospectivos, a proporção de pessoas com desfecho desfavorável em determinado grupo é uma estimativa do risco para a ocorrência desse desfecho nesse grupo.

Exemplo 7.10

Foi feito um estudo prospectivo com 1.229 gestantes de Campinas, SP, entre 2004 e 2006, para avaliar os fatores de risco comumente associados a desfechos desfavoráveis na saúde de recém-nascidos, como baixo peso ao nascer ou prematuridade. Veja os dados para um desses fatores – consumo de cigarros durante a gestação – que permitem estimar riscos.

Estimativas do risco de baixo peso ao nascer ou prematuridade segundo o consumo ou não de cigarros durante a gestação

Consumo de cigarros durante a gestação	Baixo peso ou prematuridade		Total	Risco
	Sim	Não		
Sim	44	121	165	0,2667
Não	146	918	1.064	0,1372
Total	190	1.039	1.229	

Fonte: AUDI, C.A.F. et al. Associação entre violência doméstica na gestação e peso ao nascer ou prematuridade. *J Pediatr* (Rio J) 84(1), 2008.

Risco relativo é a razão entre dois riscos. Procure dividir sempre o maior risco pelo menor para facilitar a interpretação.

O Exemplo 7.10 mostra que o risco de um nascituro ter baixo peso ao nascer ou prematuridade é maior se a mãe consumiu cigarros durante a gestação. No entanto, em quantas vezes o risco é maior? No caso do exemplo, o risco relativo é:

$$R.R. = \frac{0,2667}{0,1372} = 1,94$$

isto é, o risco de nascituro com baixo peso ao nascer ou prematuridade praticamente dobra se a mãe consumir cigarros durante a gestação.

Na área da saúde, existe muito interesse em conhecer os riscos das doenças e os fatores de risco a elas associados (fatores que aumentam o risco de uma pessoa ter determinada doença). No entanto, pelo menos à primeira vista, os estudos retrospectivos *não* permitem estimar riscos. Nesses estudos, os pesquisadores procuram pessoas com uma doença – por exemplo, úlcera gástrica – e verificam quantas delas estiveram expostas a um fator que presumem de risco – por exemplo, consumo abusivo de comida apimentada por longo tempo. Depois, procuram pessoas sem a doença e verificam quantas delas estiveram expostas ao mesmo fator para, depois, fazer comparações. Então, não é possível calcular riscos, mas apenas as proporções de pessoas que foram expostas ao fator, entre casos e controles.

No entanto – é importante saber –, quando a doença é de baixa incidência, a razão de chances tem valor próximo ao do risco relativo. Daí a importância dos estudos retrospectivos: são rápidos, baratos e dão ideia do valor do risco relativo.

Exemplo 7.11

Reveja o estudo retrospectivo para saber se câncer no pulmão ocorre mais em fumantes, apresentado no Exemplo 5.3 do Capítulo 5. Como se estima o risco de um fumante ter câncer de pulmão? Calcule a razão de chances, que é uma estimativa do risco relativo. Então, usando a razão de chances como estimativa de risco, podemos dizer que é 14 vezes mais provável um fumante ter câncer do pulmão comparativamente com um não fumante.

$$o = \frac{27 \times 647}{2 \times 622} = \frac{17469}{1244} = 14,04$$

Participantes da pesquisa segundo ter ou não câncer de pulmão e ser ou não fumante

Câncer no pulmão	Hábito de fumar		Total	Proporção de fumantes
	Sim	Não		
Presente	27	622	649	0,0416
Ausente	2	647	649	0,00308
Total	29	1.269	1.298	

Fonte: DOLL, R.; HILL, A.B. Smoking and carcinoma of the lung; preliminary report. *Br Med J* 2(4682): 739-748, 1950.

7.2. MEDIDAS DE ASSOCIAÇÃO NAS TABELAS *r* × *s*

Nas diversas áreas de ciências sociais e em epidemiologia, é bastante comum usar medidas de associação para explicar os achados de uma tabela $r \times s$. Os resultados obtidos não são, no entanto, de fácil interpretação.[98] Existem muitas medidas que têm fórmulas diferentes, propriedades diferentes e levam, portanto, a resultados diferentes. Apresentamos aqui algumas dessas medidas, todas baseadas no valor de χ2. No entanto, veja essas medidas apenas como números que *indicam* quando uma tabela mostra maior grau de associação do que outra.[99]

7.2.1. Coeficiente FI

O cálculo do coeficiente fi, que se representa pela letra grega φ (lê-se fi), é baseado no valor não corrigido de χ2, conforme visto na Seção 7.1.1. A divisão do valor de χ2 por n faz com que o valor do coeficiente não seja influenciado pelo tamanho da amostra. Por definição:

$$\varphi = \sqrt{\frac{\chi^2}{n}}$$

Exemplo 7.12

Reveja os dados apresentados no Exemplo 6.5 do Capítulo 6 e apresentados novamente aqui. O valor de χ2 foi calculado e resultou em 10,816, significante no nível de 5%. Como $n = 442$, obtemos:

$$\varphi = \sqrt{\frac{10{,}816}{442}} = 0{,}156$$

Portanto, embora o teste estatístico tenha mostrado significância, a associação entre centro e derrame recorrente é pequena.

Participantes da pesquisa segundo o centro em que foram tratados e o fato de terem ou não tido derrame recorrente

Derrame recorrente	Centro				Total
	A	B	C	D	
Sim	16	12	21	12	61
Não	179	70	78	54	381
Total	195	82	99	66	442

Alguns autores definem o coeficiente φ^2 – e não o coeficiente φ. O primeiro é, evidentemente, o quadrado do segundo. Se você estiver usando um programa de computador, verifique se é calculado o valor de φ ou o de φ^2. De qualquer forma, verifique que, para os dados do Exemplo 7.12, o valor de φ^2 é:

$$\varphi^2 = 0{,}0245$$

É melhor calcular o valor de $\$$ – e não de $\$^2$ – porque assim você obtém resultados comparáveis com os das demais medidas de associação dadas nesta seção.

7.2.2. Coeficiente de contingência de Pearson

O coeficiente de contingência de Pearson é indicado por *P* e definido pela fórmula:

$$P = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

em que *n* é o tamanho da amostra. Os coeficientes de associação variam entre 0 e 1, mas o coeficiente de contingência de Pearson não atinge o valor 1 mesmo quando a associação é perfeita. Matematicamente, seria necessário $n = 0$, o que não é possível (*n* é o tamanho da amostra).

Exemplo 7.13

Reveja os dados apresentados no Exemplo 7.12. O valor de χ2 foi calculado e resultou em 10,816. Como $n = 442$, obtemos:

$$P = \sqrt{\frac{10{,}816}{10{,}816 + 442}} = 0{,}155$$

O valor obtido para o coeficiente de contingência é pequeno (porque 0,155 está próximo de zero), indicando, novamente, pouca associação entre a ocorrência de novo derrame e o centro em que o paciente foi tratado.

7.2.3. Coeficiente de Cramér

O coeficiente de associação para tabelas $r \times s$ mais conhecido, e com melhores propriedades, é o *V* de Cramér, dado pela fórmula:

$$V = \sqrt{\frac{\frac{\chi^2}{n}}{min\{(r-1)(s-1)\}}}$$

em que min{(r – 1), (s – 1)} é o mínimo de (r – 1) e (s – 1). Em outras palavras, compare *r* e *s* e use, no denominador da fórmula, o menor deles: ou (r – 1) ou (s – 1). Note que, $s = r = 2$, o coeficiente de Cramér é igual ao coeficiente fi.

Exemplo 7.14

Reveja os dados apresentados no Exemplo 7.12. O valor de χ2 calculado resultou em 10,816 e $n = 442$. São $r = 2$ linhas e $s = 4$ colunas. Então, o mínimo de (r – 1) e (s – 1) é (r – 1) = (2 – 1) = 1. Logo:

$$V = \sqrt{\frac{\frac{10,816}{442}}{1}} = 0,156$$

O valor obtido para o coeficiente de contingência é pequeno, o que indica pouca associação entre as variáveis em análise.

7.3. COEFICIENTE DE CORRELAÇÃO DE SPEARMAN

Vimos, até aqui, medidas de associação entre duas variáveis nominais. No entanto, pode haver interesse em estudar o *grau da correlação* entre variáveis ordinais e entre variáveis numéricas. A medida de correlação mais conhecida é o coeficiente de correlação de Pearson,[100] que mede o grau de correlação entre duas variáveis numéricas.

Existe, porém, uma *alternativa não paramétrica* para o coeficiente de correlação de Pearson. É o coeficiente de correlação de Spearman, que deve ser usado quando os dados observados são ordinais ou quando as variáveis numéricas em análise não têm distribuição normal bidimensional. Para calcular o coeficiente de correlação de Spearman:

Primeiro passo: Converta os valores observados das duas variáveis em postos.

Segundo passo: Calcule a diferença D entre cada par de postos, isto é, calcule a diferença entre o posto de Y e o posto de X para a mesma unidade. Se os cálculos estiverem corretos, $\sum D$, isto é, o somatório dessas diferenças, deverá ser igual a zero.

Terceiro passo: Calcule o quadrado de cada diferença (D^2) e depois faça o somatório, isto é, calcule $\sum D^2$.

Quarto passo: Calcule o coeficiente de correlação de Spearman usando a fórmula:

$$r_s = 1 - \left[\frac{6 \sum D^2}{n\left(n^2 - 1\right)} \right]$$

Nessa fórmula, rs é o coeficiente de correlação de Spearman, D é a diferença entre os postos de uma mesma unidade e n é o número de pares de unidades da amostra.

Quinto passo: Como o coeficiente de correlação varia entre –1,00 (correlação perfeita negativa) e +1,00 (correlação perfeita positiva), você pode testar a hipótese da nulidade de que, na população, as variáveis X e Y não estão correlacionadas. Escolha uma das seguintes alternativas:

- X e Y têm correlação (hipótese bilateral, porque essa correlação tanto pode ser positiva como negativa).
- X e Y têm correlação positiva (hipótese unilateral, porque a correlação só poderá ser positiva).
- X e Y têm correlação negativa (hipótese unilateral, porque a correlação só poderá ser negativa).

Sexto passo: Compare o valor calculado de r_s com o valor crítico dado na Tabela 11 do Apêndice. Essa tabela fornece valores críticos tanto para testes unilaterais como bilaterais, para os níveis de significância $\alpha = 0{,}05$ e $\alpha = 0{,}01$. Rejeite a hipótese de que não há correlação entre as variáveis toda vez que o valor calculado de r_s for maior do que o valor crítico. Entretanto, é melhor usar um programa de computador para fazer esse teste: além de mais seguro, o programa

fornece o *p*-valor.

Quando não ocorrem empates, o coeficiente de correlação de Spearman tem o mesmo valor do coeficiente de correlação de Pearson calculado com os postos – e não com os dados originais. Se existem empates, os resultados são diferentes, mas muitas vezes a diferença é desprezível.

Exemplo 7.15

Foram obtidos escores de ansiedade por meio de um questionário (Spielberger State Trait Anxiety Inventory) e da pressão sistólica, em milímetros de mercúrio, de pacientes antes de exodontias múltiplas. Os dados estão apresentados em tabela e em diagrama de dispersão.

Escores de ansiedade e pressão sistólica

Ansiedade	**Pressão sistólica**
43,0	139,7
51,2	160,0
41,0	130,0
48,2	150,0
37,1	120,0
48,8	155,0
37,6	125,0
45,8	150,0
34,8	120,0

Fonte: OLIVEIRA, J.A.G.P.; GUIMARÃES, E.C.; OLIVEIRA, L.S. Avaliação da ansiedade e dos parâmetros cardiovasculares em pacientes hipertensos submetidos ao uso da pré-medicação Diazepam e da solução anestésica Bupivacaína (Neocaína a 0,5% sem epinefrina) em exodontias múltiplas. Estudo duplo-cego. *Rev ABO Nac* 7(2): 96-99, 1999.

Vamos optar pelo coeficiente de correlação de Spearman, uma vez que escores de ansiedade não têm distribuição normal e a amostra é pequena.

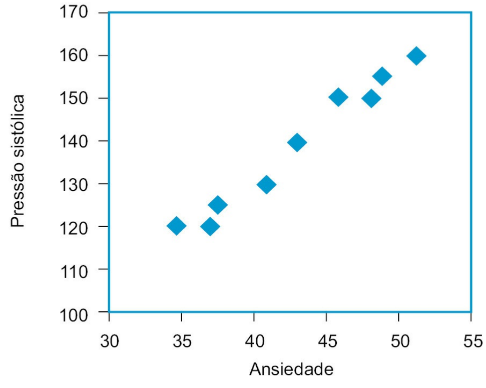


Figura 7.3. Medidas de ansiedade e pressão sistólica.

Primeiro passo: Atribua postos às medidas de ansiedade e às medidas de pressão sistólica. Como há valores iguais de pressão sistólica, houve *empates*. Atribua, aos valores iguais, posto igual à média dos postos que seriam ocupados por eles caso não fossem iguais (não houvesse empate).

Cálculos intermediários

Ansiedade	**Pressão sistólica**	**Posto para *X***	**Posto para *Y***	***D***	***D*²**
43	139,7	5	5	0	0
51,2	160	9	9	0	0
41	130	4	4	0	0
48,2	150	7	6,5	–0,5	0,25
37,1	120	2	1,5	–0,5	0,25
48,8	155	8	8	0	0
37,6	125	3	3	0	0
45,8	150	6	6,5	0,5	0,25
34,8	120	1	1,5	0,5	0,25

Segundo passo: Calcule a diferença entre cada par de postos, que indicaremos por *D*. Por exemplo, a diferença entre os postos conferidos ao quarto paciente é 6,5 – 7 = – 0,5. Não é necessário calcular ΣD, mas essa soma permite verificar se o cálculo feito anteriormente está correto, uma vez que ΣD é, necessariamente, igual a zero.

Terceiro passo: Calcule o quadrado de cada diferença (D^2) e depois some (ΣD^2). Para a Tabela 7.14, o somatório dos quadrados de *D* é:

$$\Sigma D^2 = 1$$

Quarto passo: Usando a fórmula, calcule o coeficiente de correlação de Spearman:

$$r_s = 1 - \left[\frac{6 \times 1}{9 \times \left(9^2 - 1\right)}\right] = 0{,}99167$$

Quinto passo: Para um teste bilateral com α = 0,05 e *n* = 9, a Tabela 11 do Apêndice dá o valor crítico: 0,6833. Como o valor absoluto de r_s calculado com base na amostra é maior do que o valor dado na tabela (0,99617 > 0,6833), você rejeita a hipótese da nulidade. Em outras palavras, a medida de ansiedade obtida por meio de questionário está correlacionada com a pressão sistólica.

Se você usar um programa de computador para fazer os cálculos, verá que o teste estatístico é feito pelo teste *t*. No entanto, você já sabe como tirar as conclusões. O *Statistica*, por exemplo, dará os resultados:

	Spearman Rank Order Correlations			
	Valid N	**Spearman R**	**t (n-2)**	**p-level**
Ansiedade e pressão sistólica	9	0,991632	20,32237	1,75E-07

RESUMO E OBJETIVO DO CAPÍTULO

Depois de estudar este Capítulo, você deve ser capaz de calcular e interpretar:

1. Medidas de associação para dados apresentados em tabelas 2 × 2.
2. Medidas de associação para dados apresentados em tabelas $r \times s$.
3. Coeficiente de correlação de Spearman.

---

[89] Este valor, porém, só ocorre quando as amostras são de mesmo tamanho.

[90] Veja FLEISS, JL. *Statistical methods for rates and proportions.* 2nd ed. New York: Wiley, 1981. p. 60.

[91] Dados fictícios criados com base no trabalho de VEDOVELO, M. *Prevalência de agenesias dentárias em escolares de Piracicaba*. Dissertação de mestrado. FOPUnicamp, 1972.

[92] O coeficiente ∘ também é conhecido como coeficiente de Yule.

[93] Dados fictícios com base em NARCISO DA CRUZ, C.M. *Estresse e fumo como fatores de risco para doença periodontal*. Dissertação. Universidade Camilo Castelo Branco, 2000.

[94] Doença da gengiva.

[95] Probabilidade de rejeitar a hipótese da nulidade quando a hipótese alternativa é verdadeira.

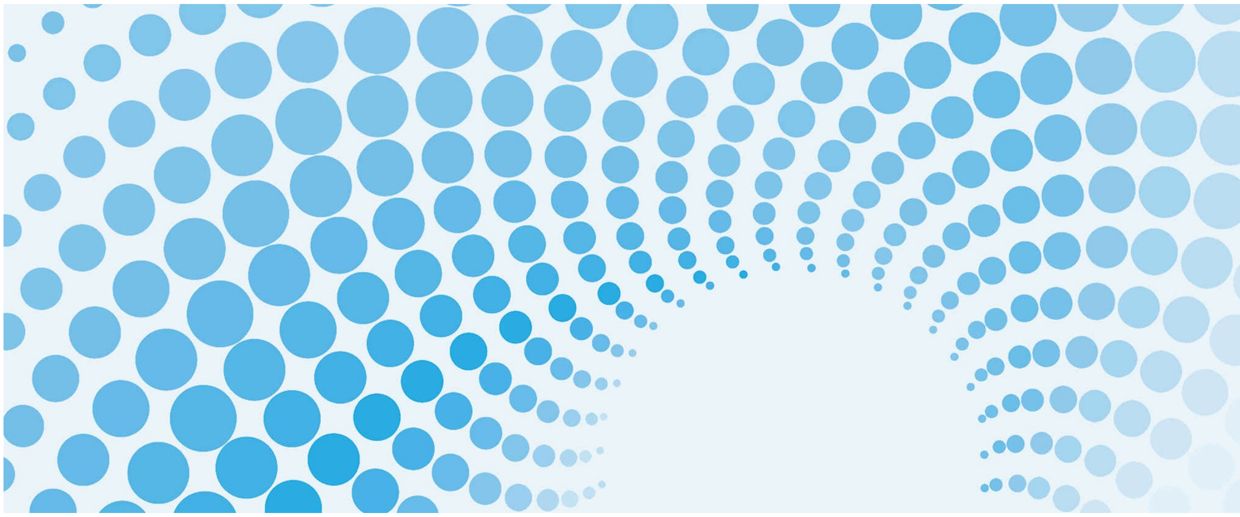
[96] Alguns autores consideram que calcular a diferença entre as chances seria mais adequado, mas a razão de chances já se popularizou entre pesquisadores da área de saúde.

[97] A razão de chances pode ter valores altos: veja o exemplo 7.11, em que a razão de chances chega a 14,04.

[98] HILDEBRAND, D.K.; LAING, J.D.; ROSENTHAL, H. *Prediction analysis of cross classifications*. Nova York: Wiley, 1977. p.49.

[99] BISHOP, Y.M.M.; FIENBERG, S.E.; HOLLAND, P.W. *Discrete multivariate analysis*: *Theory and Practice*. Cambridge, Massachussets: The Mit Press, 1977.

[100] Veja, por exemplo: VIEIRA, S. *Introdução à Bioestatística*. 4. ed. Rio de Janeiro: Campus-Elsevier, 2008.



7.4.1. Perguntou-se a 43 pessoas se elas acreditavam em horóscopos e em seres extraterrestres. Com os dados da Tabela 7.1, calcule o coeficiente $ (fi) e interprete.

Tabela 7.1 – Distribuição dos entrevistados segundo aquilo em que acreditam

Extraterrestre	**Horóscopo**	
	Sim	**Não**
Sim	14	10
Não	6	13

Fonte: http://flowjo.typepad.com/the_daily_dongle/files/Phi-coefficient.pdf

7.4.2. Com base nos dados apresentados na Tabela 7.2, faça o teste de χ2 e calcule o coeficiente de associação fi. Interprete o resultado.

Tabela 7.2 – Escolares distribuídos de acordo com o gênero e a presença ou não de hábitos orais nocivos

Sexo	**Hábitos orais nocivos**		**Total**
	Sim	**Não**	
Masculino	450	582	1.032
Feminino	535	449	984
Total	985	1.031	2.016

Fonte: SILVA FILHO, O.G. et al. Hábitos de sucção e má oclusão: epidemiologia na dentadura decídua. *R Clin Ortodon Dental Press* 2(5): 57-74, 2003.

7.4.3. Por meio de inquérito domiciliar, foram entrevistadas 200 mulheres da localidade de Puerto Leoni, Misiones, Argentina, sobre o exame Papanicolaou. As respostas sobre a escolaridade dessas mulheres e sobre o conhecimento delas a respeito do exame Papanicolaou são apresentadas na Tabela 7.3. Faça o teste de χ2 e calcule o coeficiente de associação gama.

Tabela 7.3 – Conhecimento das mulheres sobre o Papanicolaou em função da escolaridade

Escolaridade	**Conhecimento**	
	Inadequado	**Adequado**
6 anos ou menos	60	43
7 anos ou mais	39	58

Fonte: GAMARRA, C.J. et al. Conhecimentos, atitudes e prática do exame Papanicolaou entre mulheres argentinas. Rev Saúde Pública 39(2): 270-276, 2005.

7.4.4. Com base nos dados apresentados na Tabela 7.4, calcule riscos, risco relativo e faça o teste de χ2.

Tabela 7.4 – Óbito neonatal e diabetes melito

Diabetes melito	**Óbito neonatal**	
	Sim	**Não**
Sim	3	21
Não	21	830

Fonte: CUNHA, A.A.; PORTELA, M.C.; AMED, A.M.; CAMANO, L. Diabetes *mellitus* tipo 1 (insulino-dependente) e gravidez: conduta obstétrica e resultado perinatal. *Go Atual* 6(5): 24-26, 2001.

7.4.5. Para comparar a eficácia e a tolerabilidade da combinação de candesartan cilexetil e hidroclorotiazida com a combinação de losartan e hidroclorotiazida na hipertensão, foram obtidos os dados apresentados na Tabela 7.5. Por "paciente controlado" entende-se aquele que, na última consulta, tinha PAD (pressão arterial diastólica) menor que 90 mmHg na posição sentado e por "respondedor" aquele que tinha PAD maior que 90 mmHg, mas havia conseguido diminuição maior do que 10 mmHg entre a primeira e a última consulta. Por paciente "não respondedor" entende-se aquele que não conseguiu PAD menor que 90 mmHg, nem diminuição da PAD maior do que 10 mmHg entre a primeira e a última consulta. Calcule coeficientes de associação e compare os resultados.

Tabela 7.5 – Pacientes segundo a resposta ao tratamento e a combinação de drogas

Paciente	**Combinação de drogas**	
	Candesartan + hidroclorotiazida	**Losartan + hidroclorotiazida**
Controlados	85	65
Respondedores	10	10
Não respondedores	44	57

Fonte: OHMAN, K.P.; MILON, H.; VALNES, K. Eficácia e tolerabilidade do comprimido da combinação de Candesartan Cilexetil e Hidroclorotiazida na hipertensão essencial insuficientemente controlada – comparação com a combinação de Losartan e Hidroclorotiazida. *Blood Press* 9(4): 214-220, 2000.

7.4.6. Imagine que você obteve as médias de oito alunos de uma faculdade, em aulas teóricas (*X*) e em aulas práticas (*Y*). Os dados estão na Tabela 7.6. Calcule o coeficiente de correlação de Spearman.

Tabela 7.6 – Médias de oito alunos de uma faculdade em aulas teóricas (X) e em aulas práticas (Y)

Aluno	***X***	***Y***
1	30	60
2	55	55
3	65	45
4	80	90
5	85	50
6	45	85
7	50	80
8	60	95

7.4.7. São dadas[101] as notas de três juízes aos trabalhos de 10 artistas. Calcule o coeficiente de correlação de Spearman para cada par de juízes e decida: a) quais são os dois juízes que têm opiniões mais parecidas? b) quais são os dois juízes que têm opiniões mais diferentes?

Juiz A: 5; 8; 4; 2; 3; 1; 10; 7; 9; 6.

Juiz B: 3; 10; 1; 4; 2; 5; 6; 7; 8; 9

Juiz C: 8; 5; 6; 4; 10; 2; 3; 1; 7; 9.

7.4.8. Foi feito um levantamento de sintomas depressivos com estudantes de ambos os sexos com idades entre 10 e 17 anos. A avaliação foi feita por meio do questionário de autoavaliação denominado Children's Depression Inventory (CDI). Com base nos dados apresentados na Tabela 7.7, faça o teste de $\chi 2$ e calcule medidas de associação.

Tabela 7.7 – Distribuição dos participantes da pesquisa segundo o sexo e a presença de sintomas depressivos

Sexo	Sintomas depressivos	
	Sim	Não
Masculino	26	168
Feminino	68	201

Fonte: BAHLS, S.C. Epidemiology of depressive symptoms in adolescents of a public school in Curitiba, Brazil. *Rev Bras Psiquiatr* 24(2): 63-67, 2002.

7.4.9. Foi feito um questionário[102] para saber o que mais contribui para a felicidade das pessoas: amigos e vida social, trabalho e *hobby* ou condição física e saúde. Responderam ao questionário 80 pessoas solteiras, 120 casadas e 100 viúvas ou divorciadas. Os resultados estão apresentados na Tabela 7.8. Faça um teste estatístico e calcule uma medida de associação.

Tabela 7.8 – O que mais contribui para a felicidade segundo o estado civil das pessoas

Alternativa	Estado civil		
	Solteiro	Casado	Viúvo ou divorciado
Amigos e vida social	41	49	42
Trabalho e hobby	27	50	33
Saúde e condição física	12	21	25
Total	80	120	100

7.4.10. Foi feito um estudo transversal de 976 formulários de registro de dados relativos a casos atendidos pelo SOS Criança e concluídos em 1993. Desses, 587 referiam-se a denúncias de maus-tratos. As denúncias foram classificadas em confirmadas e não confirmadas, segundo o perfil do notificante. Os dados estão na Tabela 6.3 do Capítulo 6. Existe associação entre denúncia confirmada e perfil do denunciante?

---

[101] FREUND, J.E.; SMITH, R.M. *Statistics: a first course*. 4th ed. Englewood Cliffs: Prentice Hall, 1970. p. 48.

[102] FREUND, J.E.; SMITH, R.M. *Statistics: a first course*. 4th ed. Englewood Cliffs: Prentice Hall, 1970. p. 389.

8

Outras Estatísticas

Este Capítulo cobre alguns assuntos usuais na literatura da área de saúde que não foram tratados nos capítulos anteriores, como os testes diagnósticos, a concordância entre examinadores e o número necessário tratar. Esses assuntos têm, em comum, a matemática relativamente simples.

8.1. TESTES DIAGNÓSTICOS

Para entender como uma doença se transmite e para dar cuidados médicos efetivos, é preciso saber quem está doente e quem não está doente. O diagnóstico da doença – seja por exame clínico, por exame laboratorial, por imagem ou por uma combinação de exames – precisa responder à pergunta: "O paciente está doente ou não está doente?". A qualidade dos testes diagnósticos é, portanto, fundamental.

Muitas pesquisas são feitas com a finalidade de melhorar as técnicas de diagnóstico. A análise estatística dessas pesquisas não é difícil, mas a terminologia usada na área é muito específica. Vamos ver aqui os testes diagnósticos que só podem resultar em *positivo* ou *negativo*. Por exemplo, se um médico suspeita de que seu paciente esteja infectado por HIV, pede um exame de laboratório. O resultado pode ser "positivo" ou "negativo". É razoável pensar que, se o resultado do exame der *positivo*, o paciente está *infectado* e, se der *negativo*, o paciente *não está infectado.* No entanto, é preciso considerar a possibilidade de o teste ter dado resultado errado.

O médico precisa ter uma estimativa da *probabilidade de erro dos testes diagnósticos,* isto é, uma estimativa da *probabilidade* de obter um resultado positivo para quem não tem a doença e *da probabilidade* de um resultado negativo para quem tem a doença. Vamos ver a terminologia usada na área. Dizemos que o resultado do teste é:

- *Verdadeiro-positivo* (VP): quando *detecta* a doença em quem *tem* a doença.
- Falso-negativo (FN): quando *não* detecta a doença em quem *tem* a doença.
- *Verdadeiro-negativo* (VN): quando não detecta a doença em quem não *tem* a doença.
- Falso-positivo (FP): quando *detecta* a doença em quem *não* tem a doença.

Para estimar a probabilidade de ocorrerem falsos-positivos ou falsos-negativos em testes diagnósticos, são feitos *grandes levantamentos de dados*. Então, se n_1 doentes forem submetidos a determinado teste, teremos *VP* verdadeiros-positivos e *FN* falsos-negativos. Se n_2 voluntários não doentes forem submetidos a esse teste, teremos *VN* verdadeiros-negativos e *FP* falsos-positivos.

Os resultados para um teste diagnóstico dicotômico (positivo ou negativo) são apresentados em tabelas de contingência. Colocamos os resultados do teste (positivo ou negativo) nas linhas e o fato de a doença estar ou não presente nas colunas. Veja a Tabela 8.1 e a Figura 8.1.

Tabela 8.1 – Resultados do teste diagnóstico

Resultado	**Doença**	
	Sim	**Não**
Positivo	VP	FP
Negativo	FN	VN
Total	n_1	n_2

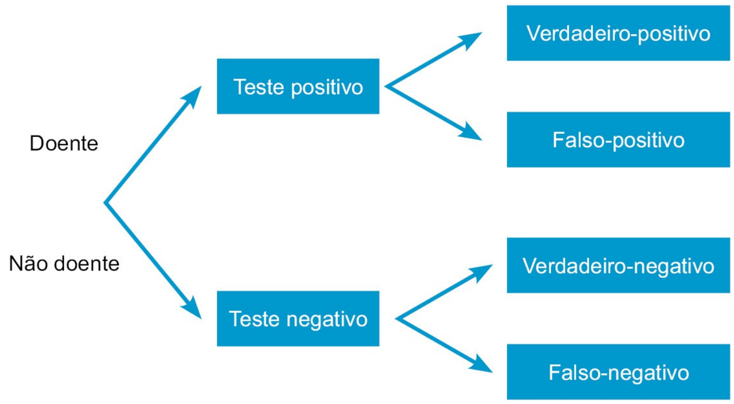


Figura 8.1 – Teste diagnóstico.

Sensibilidade (S) do teste: é a proporção de verdadeiros-positivos (resultados positivos corretos) no total de pessoas com a doença.

$$S = \frac{VP}{n_1}$$

Especificidade (E) do teste: é a proporção de verdadeiros-negativos (resultados negativos corretos) no total de pessoas sem a doença.

$$E = \frac{VN}{n_2}$$

Exemplo 8.1

Considere que um teste diagnóstico para detectar determinada doença foi aplicado em 1.000 participantes de pesquisa: 400 tinham a doença e 600 *não* tinham a doença. Os resultados do teste foram positivos em 380 doentes e negativos em 360 participantes sem a doença.

Resultados do teste diagnóstico

Resultado	**Doença**		**Total**
	Sim	**Não**	
Positivo	380	240	620
Negativo	20	360	380
Total	400	600	1.000

$$S = \frac{380}{400} = 0{,}95$$

$$E = \frac{360}{600} = 0{,}60$$

O *teste é sensível*, porque a probabilidade de dar resultado *positivo* quando a pessoa *tem* a doença é alta (acertou em 95% dos casos da amostra). Entretanto, o teste não é específico, porque a probabilidade de dar negativo em pessoas que não têm a

doença é relativamente baixa (acertou em 60% dos controles da amostra).

Qual é o problema com os testes sensíveis? *A probabilidade de ocorrerem falsos-positivos é alta*. No Exemplo 8.1, a probabilidade estimada de falsos-positivos é

$$\frac{240}{620} = 0{,}387$$

Se o teste é sensível, uma pessoa que *não* tem a doença pode receber a informação de que tem a doença. Escolha um teste sensível se

- A *doença não puder ser negligenciada*. Se a pessoa tiver a doença, o teste tem alta probabilidade de mostrar que a pessoa é doente.
- For necessário detectar pessoas doentes na população.

Exemplo 8.2

Um teste diagnóstico foi feito em 1.200 participantes de pesquisa: 500 casos (tinham a doença para a qual o teste foi proposto) e 700 controles (não tinham a doença). Os resultados do teste foram positivos em 320 casos e negativos em 630 controles.

Resultados do teste diagnóstico

Resultado	**Doença**		**Total**
	Sim	**Não**	
Positivo	320	70	390
Negativo	180	630	810
Total	500	700	1.200

$$S = \frac{320}{500} = 0{,}64$$

$$E = \frac{630}{700} = 0{,}90$$

O *teste é específico*, porque a probabilidade de dar resultado *negativo* em pessoas que *não têm* a doença é alta (acertou em 90% dos controles). No entanto, o teste não é sensível, porque a probabilidade de dar resultado positivo em pessoas que têm a doença é baixa (acertou em 64% dos casos).

Qual é o problema com os testes específicos? A *probabilidade de ocorrerem falsos-negativos é alta*. No Exemplo 8.2, a probabilidade estimada de falsos-negativos é

$$\frac{180}{500} = 0{,}36$$

Escolha um teste específico *se*

- *O diagnóstico da doença for traumático.* Para a pessoa que não tem a doença, o teste indica

isso com alta probabilidade.

- Para fechar um diagnóstico. Se o resultado for negativo, a pessoa muito provavelmente não tem a doença.

Acurácia (A) é a proporção dos resultados corretos (tanto positivos como negativos) na amostra.

$$A = \frac{VP + VN}{n}$$

A acurácia *não* é adequada para julgar um teste diagnóstico. Isso porque um valor alto de acurácia não diz se o teste tem maior probabilidade de detectar verdadeiros-positivos ou de detectar verdadeiros-negativos.

As definições de sensibilidade e especificidade também se aplicam aos sintomas e sinais. Assim:

- Se o sintoma ou sinal é *altamente sensível*, aparece em quase todos os doentes.
- Se o sintoma ou sinal é *altamente específico*, a ausência dele praticamente exclui a possibilidade de o indivíduo ter a doença.

8.1.1. Cuidados no levantamento de dados para o estudo de testes diagnósticos

Classificação de presença ou ausência da doença: deve ser feita usando como referência um "padrão-ouro" (*gold standard*) e deve ser confirmada por diferentes clínicos. O resultado dessa classificação não deve ser conhecido pelos pesquisadores que avaliam o teste.

Seleção dos participantes da pesquisa: devem ser incluídas pessoas saudáveis e pessoas com a doença que se pretende diagnosticar, mas não devem ser incluídos apenas os doentes em estado muito grave.

8.1.2. Valores preditivos

Na prática clínica, o que importa é o diagnóstico correto. Isso significa ter um teste com alta probabilidade de dar resultado positivo para o portador da doença e alta probabilidade de dar negativo para quem não tem a doença. Em outras palavras, o que interessa é saber a probabilidade de o paciente *ter a doença*, dado que o teste resultou *positivo*, e a probabilidade de o paciente *não ter a doença*, dado que o teste resultou *negativo*. Observe a Figura 8.2.

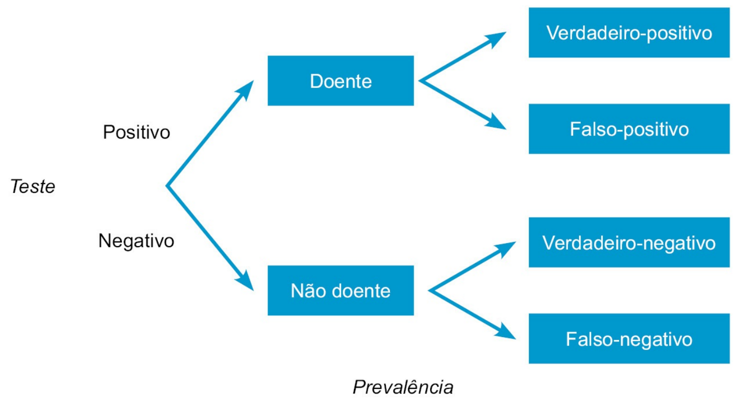


Figura 8.2 – Teste diagnóstico.

Veja, então, as definições:

Valor preditivo de um teste positivo (*VPP*): é a proporção de resultados positivos corretos no total de resultados positivos.

$$VPP = \frac{VP}{VP + FP}$$

Valor preditivo de um teste negativo (*VPN*): é a proporção de resultados negativos corretos no total de resultados negativos.

$$VPN = \frac{VN}{FN + VN}$$

Exemplo 8.3

Reveja o Exemplo 8.1: um teste diagnóstico para detectar determinada doença foi aplicado em 1.000 participantes de pesquisa: 400 tinham a doença e 600 não tinham a doença. Os resultados do teste foram positivos em 380 dos casos e negativos em 360 dos controles.

Resultados do teste diagnóstico

Resultado	Doença		Total
	Sim	**Não**	
Positivo	380	240	620
Negativo	20	360	380
Total	400	600	1.000

$$VPP = \frac{380}{380 + 240} = 0{,}613$$

$$VPN = \frac{360}{20 + 360} = 0{,}947$$

Valores preditivos são muito úteis para os clínicos, mas têm a desvantagem de depender da prevalência da doença.[103]

Prevalência é a proporção de pessoas com a doença na população. Veja, com base em exemplos,[104] como os valores preditivos mudam quando mudam as prevalências.

As Tabelas 8.2 e 8.3 apresentam dados fictícios de amostras com n = 344 indivíduos. A sensibilidade e a especificidade dos testes foram mantidas iguais.

Na Tabela 8.2 estão os dados obtidos de uma população em que a prevalência da doença é 75%. Então, espera-se que, na amostra, 0,75 × 344 = 258 indivíduos tenham a doença.

Na Tabela 8.3 estão os dados obtidos de uma população em que a prevalência da doença é 25%. Então, espera-se que, na amostra, 0,25 × 344 = 86 indivíduos tenham a doença. Veja como mudam os valores preditivos quando se mudam as prevalências.

Tabela 8.2 – Relação entre a presença ou não da doença e o resultado do teste (Prevalência = 0,75)

Resultado do teste	**Doença**		**Total**
	Presente	**Ausente**	
Positivo	231	32	263
Negativo	27	54	81
Total	258	86	344

Tabela 8.3 – Relação entre a presença ou não da doença e o resultado do teste (Prevalência = 0,25)

Resultado do teste	**Doença**		**Total**
	Presente	**Ausente**	
Positivo	77	96	173
Negativo	9	162	171
Total	86	258	344

Tabela 8.4 – Análise das propriedades dos exames para diagnóstico apresentados nas Tabelas 8.2 e 8.3

Propriedades	**Prevalência**	
	0,75	**0,25**
Sensibilidade	0,895	0,895
Especificidade	0,628	0,628
Valor preditivo do teste positivo	0,878	0,445
Valor preditivo do teste negativo	0,667	0,947

8.1.3. Razão de verossimilhanças

Além das maneiras já descritas, há outra forma de avaliar o desempenho de um teste diagnóstico. É pela *razão de verossimilhanças*, que significa razão entre duas probabilidades.

Razão de verossimilhanças é a razão entre a probabilidade de resultados positivos nas pessoas que têm a doença e a probabilidade de resultados positivos em quem não tem a doença.

$$Razão = \frac{\frac{VP}{VP + FN}}{\frac{FP}{FP + VN}}$$

$$Razão = \frac{Sensibilidade}{1 - Especificidade}$$

A razão de verossimilhança pode ser vista como indicadora do valor do teste para aumentar a certeza sobre diagnósticos positivos. Uma razão de verossimilhança maior que 1 indica que o resultado do teste está associado à presença da doença, enquanto uma razão de verossimilhança inferior a 1 indica que o resultado do teste está associado à ausência de doença.

Para os dados das Tabelas 8.2 e 8.3, a estimativa da razão de verossimilhança indica que o teste está associado à presença da doença, porque é:

$$Razão = \frac{Sensibilidade}{1 - Especificidade} = \frac{0{,}895}{1 - 0{,}628} = 2{,}41$$

8.2. NÚMERO NECESSÁRIO TRATAR (NNT)

O *NNT* ou *número necessário tratar* (*number needed to treat*) é o número de pacientes que precisam receber uma intervenção específica para que mais um deles se beneficie dentro do período de tempo estipulado para o ensaio.

Para calcular o *NNT*, existe uma fórmula.[105] Então, seja p_t a proporção de sucessos no grupo tratado e p_c a proporção de sucessos no grupo controle, como mostrado na Tabela 8.5.

Tabela 8.5 – Proporções de sucessos e fracassos nos dois grupos, tratado e controle

Evento	**Grupo**	
	Tratado	**Controle**
Sucesso	p_t	p_t
Fracasso	$1-p_t$	$1-p_c$
Total	1	1

A fórmula para calcular o número necessário tratar é

$$NNT = \frac{1}{p_t - p_c}$$

Note que 1- p_t é o risco de fracasso no grupo tratado e 1- p_c é o risco de fracasso no grupo controle.

Exemplo 8.4

Imagine que, para verificar se uma nova intervenção aumenta a probabilidade de sobrevida decorridos 5 anos após o diagnóstico de câncer no pulmão estágio III, foi feito um ensaio clínico randomizado com 450 pacientes: 200 foram designados para o controle e 250 para a nova intervenção. Morreram 160 pacientes do grupo controle e 175 pacientes do grupo submetido à nova intervenção no período de tempo estudado, como mostrado na tabela.

Proporção de sobreviventes segundo o grupo

Evento	**Grupo**	
	Tratado	**Controle**
Sucesso	75	40
Fracasso	175	160
Total	250	200
Proporção (sobreviventes)	$p_t = \frac{75}{250} = 0,3$	$p_t = \frac{40}{200} = 0,2$

$$NNT = \frac{1}{p_t - p_c} = \frac{1}{0,3 - 0,2} = 10$$

No exemplo, $NNT = 10$ é o número de pacientes que precisam ser tratados para que, dentro de 5 anos, ocorra o desfecho desejado (sobrevida) em um paciente a mais do que o esperado para o grupo controle. Talvez seja difícil entender. Vamos então construir tabelas consecutivamente, mantendo as proporções de sobreviventes do Exemplo 8.4, mas simplificando os números.

Proporção de sobreviventes segundo o grupo (grupos de mesmo tamanho)

Evento	Grupo	
	Tratado	**Controle**
Sucesso	60	40
Fracasso	140	160
Total	200	200
Proporção (sobreviventes)	$p_t = \frac{60}{200} = 0{,}3$	$p_t = \frac{40}{200} = 0{,}2$

Note que todos os números na tabela que acabamos de ver são divisíveis por 20. Vamos, então, fazer essa simplificação.

Proporção de sobreviventes segundo os grupos (grupos de tamanho 10)

Evento	Grupo	
	Tratado	**Controle**
Sucesso	3	2
Fracasso	7	8
Total	10	10
Proporção (sobreviventes)	$p_t = \frac{3}{10} = 0{,}3$	$p_t = \frac{2}{10} = 0{,}2$

Veja a explicação do $NNT = 10$: se forem tratados 10 pacientes, espera-se que sobrevivam 3, 1 a mais do que os 2 que sobreviveriam no grupo controle. Portanto:

- $NNT = 100$ significa que 100 pacientes precisam ser tratados para que mais 1 deles tenha o efeito desejado.
- $NNT = 10$ significa que 10 pacientes precisam ser tratados para que mais 1 deles tenha o efeito desejado.
- $NNT = 1$ significa que um paciente precisa ser tratado para que tenha o efeito desejado.

Então, fica evidente: quanto menor for o NNT, maior é o benefício da terapia. $NNT = 1$ é, portanto, o ideal, porque significa que todo paciente tratado é beneficiado.

Exemplo 8.5

Imagine que 12% das pessoas com determinado diagnóstico morrem dentro de 1 ano. Uma nova droga reduziu a mortalidade para 8%. Como se calcula o NNT?

$$p_t = 1 - 0{,}08 = 0{,}92$$

$$p_c = 1 - 0{,}12 = 0{,}88$$

$$NNT = \frac{1}{0{,}92 - 0{,}88} = \frac{1}{0{,}04} = 25$$

Em média, 25 pacientes teriam que receber a nova intervenção (em lugar do controle) para que se espere registrar a sobrevivência de 1 paciente a mais.

O número necessário tratar é o inverso da *redução absoluta de risco* (RAR) (*absolute risk reduction*). É muitas vezes expressa como porcentagem. Veja:

$$RAR = \textit{taxa do evento no controle} - \textit{taxa do evento no tratado}$$

Exemplo 8.6

Reveja o Exemplo 8.5. A redução absoluta de risco (*RAR*) é dada por

$$RAR = 12\% - 8\% = 4\%$$

A redução do risco de morte, devido ao tratamento, foi de 12% – 8%= 4%.

A redução absoluta de risco é uma forma útil de apresentar o resultado da pesquisa, principalmente se o leitor for um tomador de decisão.

8.3. CONCORDÂNCIA ENTRE EXAMINADORES

Nas pesquisas em que é preciso classificar as unidades em várias categorias, podem surgir discordâncias entre os examinadores sobre algumas classificações. Como saber se os examinadores concordam na maior parte das vezes? Vamos discutir o caso mais simples, de dois examinadores que classificam as unidades do estudo em várias categorias.

Veja a Tabela 8.6. Dois examinadores, A e B, classificaram a gravidade da doença de *n* pacientes em três categorias, leve, moderada e severa. As células que constituem a diagonal principal foram coloridas. Ali estão as concordâncias. Somando o valor dessas células e dividindo pelo total, você tem a porcentagem de concordâncias no total de julgamentos.

Tabela 8.6 – Gravidade da doença julgada por dois examinadores em n pacientes

Examinador A	**Examinador B**			
	Leve	**Moderada**	**Severa**	**Total**
Leve				
Moderada				
Severa				
Total				*n*

Exemplo 8.7

Dois radiologistas classificaram 85 mamografias nas seguintes categorias: “Normal”, “Doença benigna”, “Suspeita de câncer” ou “Câncer”. É fácil contar, na tabela, o número de casos em que eles estavam de acordo (em negrito).

Classificação das mamografias por dois radiologistas

Radiologista A	**Radiologista B**				**Total**
	Normal	**Doença benigna**	**Suspeita de câncer**	**Câncer**	
Normal	**21**	12	0	0	33
Doença benigna	4	**17**	1	0	22
Suspeita de câncer	3	9	**15**	2	29
Câncer	0	0	0	**1**	1
Total	28	38	16	3	85

Uma medida da concordância é dada por:

$$\frac{21+17+15+1}{85} = \frac{54}{85} = 0,635$$

No Exemplo 8.7, que acabamos de ver, *não* foi colocada uma hipótese em teste, *nem foi*

medida a associação entre as respostas – o que se obteve foi o *grau de concordância* dos radiologistas – que, no caso, foi estabelecida em 63,5%. Essa estatística é citada em muitos trabalhos. No entanto, é lógico esperar que algumas concordâncias ocorram por simples acaso – o que minimiza a importância desse resultado.

Nem todos os estatísticos concordam que se deva procurar outra estatística para indicar o grau de concordância entre examinadores. No entanto, o fato é que a estatística *kappa*, que se representa pela letra grega $\mp$ (que se lê “capa”), é bem conhecida e está em programas de computador, o que garante seu uso.

Vejamos como se calcula esse coeficiente de concordância. Primeiro, considere: se todos os examinadores concordarem com a categorização das unidades, a estatística que mede concordância deve ter valor 1. Seja I_0 a proporção das concordâncias observadas na amostra. Como já calculamos, no caso do Exemplo 8.6:

$$I_0 = \frac{21+17+15+1}{85} = \frac{54}{85} = 0,635$$

Vamos agora calcular a proporção de concordâncias que seriam obtidas para esse exemplo por simples acaso, que indicaremos por I_E. Lembre-se do que foi visto no Capítulo 6 para tabelas de contingência: sob a hipótese de que não existe associação (hipótese da nulidade), a *frequência esperada* em determinada célula é dada pelo *total da coluna* (em que a célula está) multiplicado pelo *total da linha* (em que a célula está), *dividido pelo tamanho da amostra*. Calculam-se, assim, as frequências esperadas na diagonal principal – ou seja, as frequências esperadas sob a hipótese de que as concordâncias entre examinadores seriam apenas casuais.

Radiologista A		**Radiologista B**		
	Normal	**Doença benigna**	**Suspeita de câncer**	**Câncer**
Normal	$\frac{33 \times 28}{85} = 10,87$			
Doença benigna		$\frac{22 \times 38}{85} = 9,84$		
Suspeita de câncer			$\frac{29 \times 16}{85} = 5,46$	
Câncer				$\frac{1 \times 3}{85} = 0,04$

A proporção de concordâncias que seriam obtidas por simples acaso é dada por:

$$\frac{10,87 + 9,84 + 5,46 + 0,04}{85} = \frac{26,2}{85} = 0,308$$

No Exemplo 8.7, a proporção de concordâncias observadas na amostra foi de 0,635, que indicamos por I_0. A proporção de concordâncias que ocorreriam por acaso é de 0,308, que

indicamos por I_E. Sabemos, ainda, que o máximo possível de concordâncias é 1,000. O coeficiente de concordância *kappa* é construído da seguinte forma:

- A proporção máxima de concordâncias é 1.
- Descontando, da proporção máxima, a proporção que ocorreria por acaso, tem-se $1 - I_E$.
- Descontando, da proporção observada, a proporção que ocorreria por acaso, tem-se $I_0 - I_E$.
- A estimativa do coeficiente de concordância $\mp$ é dada pela proporção.

$$\kappa = \frac{I_0 - I_E}{1 - I_E}$$

Exemplo 8.8

Para os dados do Exemplo 8.7, o coeficiente de concordância *kappa* é:

$$\kappa = \frac{0,635 - 0,308}{1 - 0,308} = 0,472$$

Se a concordância entre examinadores for perfeita, o coeficiente de concordância $\mp$ tem valor máximo, que é 1. Se não houver concordância entre os examinadores, o valor de $\mp$ é zero. No entanto, como interpretar valores entre 0 e 1? Você pode seguir a recomendaçãoda Tabela 8.7.[106]

Tabela 8.7 – Classificação da força da concordância entre examinadores em função do valor de kappa

Valor de $\mp$	**Força da concordância**
Menor do que 0,21	Pobre
0,21 $\vdash$ 0,41	Regular
0,41 $\vdash$ 0,61	Moderada
0,61 $\vdash$ 0,81	Boa
0,81 $\vdash$ 1,00	Muito boa

Os programas para computador calculam o valor de *kappa* e dão o intervalo de confiança. A fórmula é complicada, mas você a encontra em livros de estatística mais avançados.[107]

8.4. ALFA DE CRONBACH

A confiabilidade de um instrumento de medida tem diferentes aspectos. Existem, pois, diferentes estatísticas para estimar confiabilidade, cada qual avaliando um aspecto da conformidade do instrumento. Podem ser avaliadas:

- A *confiabilidade entre examinadores*, ou seja, o grau com que diferentes examinadores veem o mesmo fenômeno, usando o mesmo instrumento.
- A *confiabilidade do teste-reteste*, isto é, a consistência das medidas feitas com o mesmo instrumento de medida em ocasiões diferentes.
- A *confiabilidade de forma paralela*, que é a consistência dos resultados de dois instrumentos diferentes, mas construídos da mesma maneira.

Todas essas características dos sistemas de medição são estudadas em estatística para qualidade. Os nomes usados em engenharia são diferentes, mas, basicamente, os conceitos são os mesmos. Nas áreas de ciências sociais, em que são feitos testes e questionários, também se define:

- *Consistência interna de um teste ou um questionário* é a extensão em que os itens que o compõem medem o mesmo conceito ou construto. Por exemplo, se 10 questões foram projetadas para medir o mesmo construto, o respondente deveria ter coerência nas respostas. A consistência interna é, portanto, uma das quatro classes de estimativas de confiabilidade, sendo específica para testes e questionários.

Para medir a consistência de um questionário, a estatística mais usada é o *coeficiente alfa*. É fácil calcular esse coeficiente e essa estatística ainda tem a vantagem de poder ser calculada mesmo quando o questionário é aplicado uma única vez. No entanto, o coeficiente alfa nem sempre é bem interpretado.

Vamos lembrar um pouco do vocabulário da área. Todo questionário é constituído por várias perguntas, que aqui serão chamadas de itens. Depois de pronto, o questionário deve ser entregue para um grupo de pessoas selecionadas de acordo com determinado critério, para que o respondam e devolvam ao pesquisador.

As opções de resposta para cada item podem ser dicotômicas, como “Sim” e “Não”, ou escalonadas, como “Concordo plenamente”, “Concordo”, “Não concordo nem discordo”, “Discordo”, “Discordo completamente”. De qualquer forma, para o cálculo do coeficiente alfa, toda resposta deve ser transformada em postos.

Seja x_{ij} o i-ésimo posto do j-ésimo respondente, $i = 1, 2,..,k$, e $j = 1, 2,...,n$. Considere um questionário com k itens, respondido por n pessoas. Para calcular o coeficiente alfa de Cronbach, aplica-se a fórmula:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} S_i^2}{S_{soma}^2}\right)$$

k é o número de itens, n é o número de respondentes.

s^2_i é a variância dos n escores das pessoas a i-ésimo item $(i = 1, ..., k)$,

s^2_{soma} é a variância dos totais Tj ($j = 1, 2,...,n$). de escores de cada respondente.

As variâncias são calculadas pela fórmula:

$$S^2 = \frac{\sum\left(x - \bar{x}\right)^2}{n}$$

Exemplo 8.9

É dado um exemplo fictício de resultados de um questionário. Eram $k = 5$ perguntas, que só podiam ser respondidas como “Sim” ou “Não”. As respostas configuram uma variável dicotômica que precisa ser transformada em número. Então “Sim” ficou sendo 1 e “Não” ficou sendo zero. Responderam o questionário $n = 10$ pessoas.

Resultados da aplicação de um questionário com 5 itens e 10 respondentes

Respondente	**Item**					**Total de escores de cada respondente (T)**
	1º	**2º**	**3º**	**4º**	**5º**	
1	1	1	1	1	1	5
2	1	1	1	1	0	4
3	1	1	1	0	1	4
4	1	1	0	1	0	3
5	1	1	1	0	0	3
6	1	1	0	0	1	3
7	1	1	0	0	0	2
8	0	1	1	0	0	2
9	1	0	1	0	0	2
10	1	0	0	0	0	1
Total de escores de cada item	9	8	6	3	3	29

Para obter alfa, é preciso calcular as variâncias que estão apresentadas no rodapé da tabela.

Variâncias usadas na fórmula de alfa

Respondente	**Item**					**Total de escores de cada respondente (T)**
	1º	**2º**	**3º**	**4º**	**5º**	
1	1	1	1	1	1	5
2	1	1	1	1	0	4

3	1	1	1	0	1	4
4	1	1	0	1	0	3
5	1	1	1	0	0	3
6	1	1	0	0	1	3
7	1	1	0	0	0	2
8	0	1	1	0	0	2
9	1	0	1	0	0	2
10	1	0	0	0	0	1
Variância	0,1000	0,1778	0,2667	0,2333	0,2333	1,4333

Número de itens = 5

Soma das variâncias dos itens

$$0{,}1000 + 0{,}1778 + 0{,}2667 + 0{,}2333 + 0{,}2333 = 1{,}0111$$

Variância dos totais dos escores de cada respondente =1,4333

$$\alpha = \frac{5}{5-1}\left[1 - \frac{0{,}1000 + 0{,}1778 + 0{,}2667 + 0{,}2333}{1{,}4333}\right] = 0{,}36$$

A interpretação do coeficiente alfa de Cronbach é aparentemente intuitiva porque, na maior parte das vezes, os valores variam entre 0 e 1. Entende-se, então, que a consistência interna de um questionário é tanto maior quanto mais perto de 1 estiver o valor da estatística. Há muita discussão sobre os valores aceitáveis de alfa: em geral, variam entre 0,70 e 0,95.

A maneira prática de julgar o valor de alfa é comparar o valor calculado com o valor preconizado por diferentes autores em tabelas apresentadas na literatura. A regra é imprecisa, mas serve como primeira aproximação, desde que se tenha a precaução de levar em conta as limitações dessa estatística. Veja as Tabelas 8.8. e 8.9.

Tabela 8.8 – Consistência interna do questionário

Valor de alfa	**Consistência interna**
Maior que 0,80	Quase perfeita
De 0,80 a 0,61	Substancial
De 0,60 a 0,41	Moderada
De 0,41 a 0,21	Razoável
Menor que 0,21	Pequena

Fonte: Landis, J. R.; Kock, G.G. The measurement of observer agreement for categorical data. *Biometrics* 33:159, 1977.

Tabela 8.9 – Consistência interna do questionário

Valor de alfa	Consistência interna
0,91 ou mais	Excelente
0,90 ├ 0,81	Bom
0,81 ├ 0,71	Aceitável
0,71 ├ 0,61	Questionável
0,61 ├ 0,51	Pobre
Menor que 0,51	Inaceitável

Fonte: GEORGE, D.; MALLERY, P. SPSS for Windows step by step: a simple guide and reference.4th ed. Boston: Alley & Bacon, 2003. Apud GLIEM, J.A.; GLIEM, R.R. Calculating, interpreting and reporting Cronbach's alpha reliability for Likert-type scales. Disponível em: htpps: //scholarworks.iupui.edu/bitstream/handle. Acesso em outubro de 2013.

É importante saber, ao julgar o valor calculado de alfa, que:

- O *número de questões* afeta o valor de alfa. Questionários muito longos aumentam o valor de alfa, sem que isso signifique aumento de consistência interna; um valor baixo de alfa pode significar apenas um número pequeno de questões.
- A *redundância*, isto é, questões verbalizadas de forma diferente, mas praticamente iguais, aumentam o valor de alfa.
- *Correlações entre os itens do questionário* aumentam o valor de alfa. Se vários itens do questionário exibem correlações entre si, o valor de alfa aumenta. Como essas correlações são maiores quando os itens do questionário medem o mesmo construto, o pesquisador conclui que o questionário tem consistência interna, ou seja, o valor alto do coeficiente alfa de Cronbach estaria indicando o grau em que os itens medem o mesmo construto. No entanto, é preciso cuidado: pode haver uma terceira variável afetando as respostas de dois itens. Uma boa discussão ajuda muito.

Para o exemplo dado, o valor do coeficiente alfa de Cronbach é $a = 0,36$. Se os dados fossem reais (e não criados apenas para mostrar como se fazem os cálculos), o valor de alfa poderia ser explicado pelo número pequeno de perguntas e de respondentes, mas também poderia indicar que os itens (perguntas) do questionário não estavam medindo o mesmo construto ou a mesma dimensão (unidimensional).

O coeficiente alfa de Cronbach pode ser calculado a partir de programas estatísticos como SPSS (*Statistical Software for Social Sciences*) ou SAS (*Statistical Analysis System*). Esses programas fornecem análise descritiva inicial completa das respostas obtidas do questionário, bem como listagem completa da análise da confiabilidade. O coeficiente alfa de Cronbach pode ser calculado por outros programas.

RESUMO E OBJETIVO DO CAPÍTULO

Depois de estudar este Capítulo, você deverá ser capaz de calcular e interpretar as seguintes estatísticas:

- Sensibilidade e especificidade nos testes diagnósticos.
- Valores preditivos nos testes diagnósticos.
- Razão de verossimilhanças nos testes diagnósticos.
- Número necessário tratar.
- Coeficiente de concordância.
- Alfa de Cronbach.

---

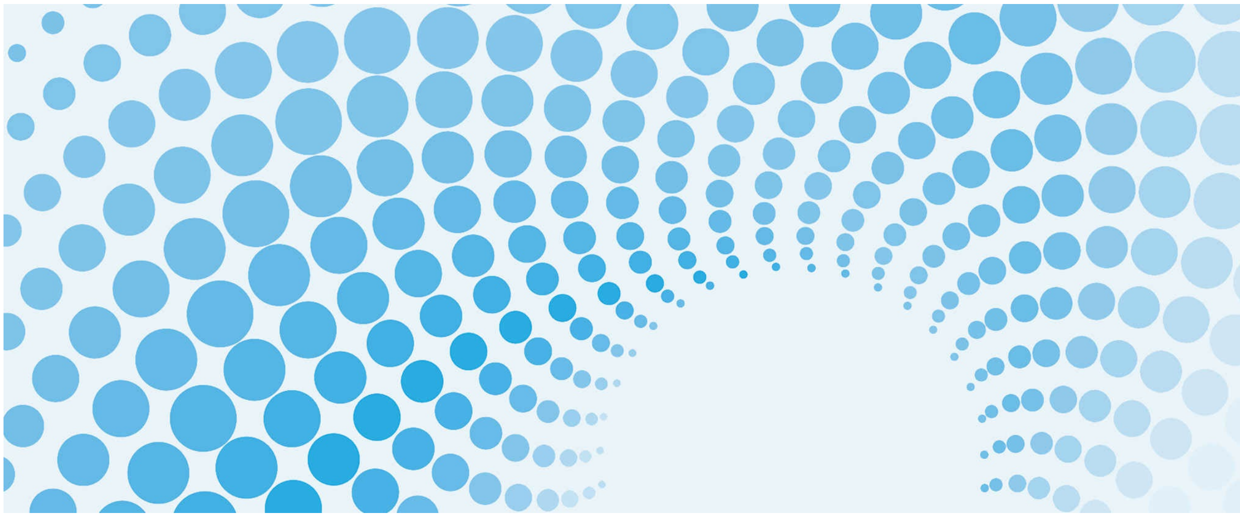
[103] Teorema de Bayes: The Harvard School Test. Disponível em: soniavieira.blogspot.com/2015/09/teorema-de-bayes-iii-harvard-medical.html. Acesso em 18 de novembro de 2017.

[104] ALTMAN, D.O. *Practical Statistics for Medical Research.* 2nd ed. London: Chapman & Hall, 1993. p. 410-413.

[105] Number Needed to Treat (NNT) Calculator – ClinCalc. Disponível em: clincalc.com/Stats/NNT.aspx

[106] ALTMAN, D.O. *Practical Statistics for Medical Research.* 2nd ed. London: Chapman & Hall, 1993. p. 404.

[107] FLEISS, J.L. *Statistical methods for rates and proportions.* 2nd ed. New York: Wiley, 1981. p. 219-220.



8.5.1. Pesquisadores analisaram o fator reumatoide (FR) no soro de 176 pacientes com artrite reumatoide e 274 sem a doença. Os dados estão apresentados na Tabela 8.10. Calcule a sensibilidade (*S*), a especificidade (*E*) e a acurácia (*A*) do teste.

Tabela 8.10 – Exame para fator reumatoide (FR) no soro de pacientes com e sem artrite reumatoide

Resultado do exame	Doença	
	Presente	Ausente
Positivo	120	24
Negativo	56	250

Fonte: MEDEIROS, M.M.C.; FERRAZ, M.B. Estudos sobre diagnósticos. *Rev Bras Reumatol* 38(4): 239-375, 1998.

8.5.2. Com os dados da Tabela 8.10, calcule o valor preditivo de teste positivo (VPP) e o valor preditivo de teste negativo (VPN).

8.5.3. Um ensaio randomizado placebo-controlado foi conduzido com participantes com hipertensão leve. Entre 100 alocados para tratamento ativo, houve 1 acidente vascular cerebral. Entre 100 alocados ao placebo, houve 2 acidentes vasculares cerebrais. Qual é o número necessário tratar (*NNT*) para evitar um único acidente vascular cerebral sob as condições deste teste?

8.5.4. Obtenha a sensibilidade, a especificidade, os valores preditivos e a precisão do exame físico prostático (EFP) no diagnóstico do adenocarcinoma de próstata (ACP). Foram levantados os dados apresentados na Tabela 8.11. Nota: a biópsia dá a situação verdadeira.

Tabela 8.11 – Distribuição dos pacientes segundo os achados no exame físico prostático e na biópsia

Exame físico prostático	Biópsia	
	Positiva	Negativa
Positivo	30	28
Negativo	31	108

Fonte: ANTOPOULOS, I.M. et al. Results of prostate cancer screening in no-symptomatic men. Disponível em: http://braziurol.com.br2001/Antonopoulos 227 234.htm

8.5.5. Para estudar a sensibilidade, a especificidade, os valores preditivos e a precisão da dosagem de antígeno prostático específico (PSA) total (valor alto aponta presença da doença) no diagnóstico do adenocarcinoma de próstata (ACP), foram levantados os dados apresentados na Tabela 8.12. Faça os cálculos. Nota: Biópsia indica a verdade, PSA total < 4 ng/mL indica resultado negativo e PSA total > 4 ng/mL indica resultado positivo.

Tabela 8.12 – Distribuição dos pacientes segundo o PSA total e a biópsia

PSA total	Biópsia	
	Positiva	Negativa
Menor do que 4 ng/mL	10	48
Igual ou maior do que 4 ng/mL	15	124

Fonte: ANTOPOULOS, I.M. et al. Results of prostate cancer screening in no-symptomatic men. Disponível em : http://braziurol.com.br 2001/Antonopoulos 227 234.htm

8.5.6. Pesquisadores analisaram as características do anticorpo antiestrato córneo (AEC) no soro de 176 pacientes com artrite reumatoide e 274 sem a doença. Os dados apresentados na Tabela 8.13 permitem calcular a razão de verossimilhanças.

Compare o resultado obtido com o resultado do Exemplo 8.7.

Tabela 8.13 – Características do anticorpo antiestrato córneo (AEC) no soro de pacientes com e sem artrite reumatoide

Resultado do exame	Doença		Total
	Presente	Ausente	
Positivo	71	4	75
Negativo	105	270	375
Total	176	274	450

Fonte: MEDEIROS, M.M.C.; FERRAZ, M.B. Estudos sobre diagnósticos. *Rev Bras Reumatol* 38(4): 239-375, 1998.

8.5.7. Um teste[108] tem sensibilidade de 99% e especificidade de 92%. Calcule o valor preditivo, tanto positivo como negativo, em uma população em que a prevalência da doença é de 5%.

8.5.8. Dois professores de um curso de pós-graduação avaliaram 20 alunos. Os critérios de avaliação eram A, B, C e D. Os critérios atribuídos estão na Tabela 8.14. Calcule *kappa* e interprete o resultado.

Tabela 8.14 – Conceitos atribuídos a 20 alunos por dois professores

Nº do aluno	Professor	
	1	2
1	A	B
2	A	A
3	A	A
4	A	A
5	A	B
6	A	B
7	A	B
8	B	C
9	A	B
10	B	A
11	A	A
12	A	A
13	A	A
14	A	A
15	A	A
16	B	C
17	B	C
18	B	B
19	A	B
20	B	A

8.5.9. Quando o valor de *kappa* é igual a 1?

8.5.10. Quando a sensibilidade e a especificidade de um exame para diagnóstico são iguais a 1?

---

[108] MOTULSKY, H. *Intuitive Biostatistics*. New York: Oxford University Press, 1995. p.139.

Apêndice

Tabela 1. Distribuição normal reduzida P(0 < *Z* < *z*)

	Último dígito									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4658	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Fonte: FREUND, J.E.; SMITH, R.N. *Statistic: A First Course*. Englewood Cliffs, Prentice Hall. p. 507, 1986.

Tabela 2. Valores de χ^2, segundo os graus de liberdade e o valor de α

Graus de liberdade	**Valor de** a		
	10%	**5%**	**1%**
1	2,71	3,84	6,64
2	4,60	5,99	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,72
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,68	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18	25,99	28,87	34,80
19	27,20	30,14	36,19
20	28,41	31,41	37,57
21	29,62	32,67	38,93
22	30,81	33,92	40,29
23	32,01	35,17	41,64
24	33,20	36,42	42,98
25	34,38	37,65	44,31
26	35,56	38,88	45,64
27	36,74	40,11	46,96
28	37,92	41,34	48,28
29	39,09	42,56	49,59
30	40,26	43,77	50,89

Fonte: SNEDECOR, G.W.; COCHRAN, W.G. *Statistical Methods*. Ames, The Jowa State University Press. p. 550-1, 1972.

Tabela 3. Valores críticos para $\sum R_1$ no teste de Mann-Whitney

Atenção: No cabeçalho estão os valores de α para um teste unilateral. Se o teste for bilateral, leia os valores que têm, no cabeçalho, o valor $\alpha/2$. Por exemplo, se $\alpha = 0{,}05$, procure o valor na coluna de $\alpha = 0{,}025$.

$n_1 = 3$

n_2	**0,005**		**0,025**	**0,05**
3				6-15
4				6-18
5			6-21	7-20
6			7-23	8-22
7			7-26	8-25
8			8-28	9-27
9	6-33		8-31	10-29
10	6-36		9-33	10-32
11	6-39		9-36	11-34
12	7-41		10-38	11-37
13	7-44		10-41	12-39
14	7-47		11-43	13-41
15	8-49		11-46	13-44

$n_1 = 4$

n_2	**0,005**	**0,01**	**0,025**	**0,05**
4			10-26	11-25
5		10-30	11-29	12-28
6	10-34	11-33	12-32	13-31
7	10-38	11-37	13-35	14-34
8	11-41	12-40	14-48	15-37
9	11-45	13-43	14-42	16-40
10	12-48	13-47	15-45	17-43
11	12-52	14-50	16-48	18-46
12	13-55	15-53	17-51	19-49
13	13-59	15-57	18-54	20-52
14	14-62	16-60	19-57	21-55
15	15-65	17-63	20-60	22-58

$n_1 = 5$

n_2	**0,005**	**0,01**	**0,025**	**0,05**
5	15-40	16-39	17-38	19-36
6	16-44	17-43	18-42	20-40
7	16-49	18-47	20-45	21-44
8	17-53	19-51	21-49	23-47
9	18-57	20-55	22-53	24-51
10	19-61	21-59	23-57	26-54
11	20-65	22-63	24-61	27-58
12	21-69	23-67	26-64	28-62
13	22-73	24-71	27-68	30-65
14	22-78	25-75	28-72	31-69
15	23-82	26-79	29-76	33-72

$n_1 = 6$

n_2	**0,005**	**0,01**	**0,025**	**0,05**
6	23-55	24-54	26-52	28-50
7	24-60	25-59	27-57	29-55
8	25-65	27-63	29-61	31-59
9	26-70	28-68	31-65	33-63
10	27-75	29-73	32-70	35-67
11	28-80	30-78	34-74	37-71
12	30-84	32-82	35-79	38-76
13	31-89	33-87	37-83	40-80
14	32-94	34-92	38-88	42-84
15	33-99	36-96	40-92	44-88

Continua

Tabela 3. Continuação

$n_1 = 7$					$n_1 = 8$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**	n_2	**0,005**	**0,01**	**0,025**	**0,05**
7	32-73	34-71	36-69	39-66	8	43-93	45-91	49-87	51-85
8	34-78	35-77	38-74	41-71	9	45-99	47-97	51-93	54-90
9	35-84	37-82	40-79	43-76	10	47-105	49-103	53-99	56-96
10	37-89	39-87	42-84	45-81	11	49-111	51-109	55-105	59-101
11	38-95	40-93	44-89	47-86	12	51-117	53-115	58-110	62-106
12	40-100	42-98	46-94	49-91	13	53-123	56-120	60-116	64-112
13	41-106	44-103	48-99	52-95	14	54-130	58-126	62-122	67-117
14	43-111	45-109	50-104	54-100	15	56-136	60-132	65-127	69-123
15	44-117	47-114	52-109	56-105					

$n_1 = 9$					$n_1 = 10$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**	n_2	**0,005**	**0,01**	**0,025**	**0,05**
9	56-115	59-112	62-109	66-105	10	71-139	74-136	78-132	82-128
10	58-122	61-119	65-115	69-111	11	73-147	77-143	81-139	86-134
11	61-128	63-126	68-121	72-117	12	76-154	79-151	84-146	89-141
12	63-135	66-132	71-127	75-123	13	79-161	82-158	88-152	92-148
13	65-142	68-139	73-134	78-129	14	81-169	85-165	91-159	96-154
14	67-149	71-145	76-140	81-135	15	84-176	88-172	94-166	99-161
15	69-156	73-152	79-146	84-141					

Continua

Tabela 3. Continuação

$n_1 = 11$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**
11	87-166	91-162	96-157	100-153
12	90-174	94-170	99-165	104-160
13	93-182	97-178	103-172	108-167
14	96-190	100-186	106-180	112-174
15	99-198	103-194	110-187	116-181

$n_1 = 12$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**
12	105-195	109-191	115-185	120-180
13	109-203	113-199	119-193	125-187
14	112-212	116-208	123-201	129-195
15	115-221	120-216	127-209	133-203

$n_1 = 13$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**
13	125-226	130-221	136-215	142-209
14	129-235	134-230	141-223	147-217
15	133-244	138-239	145-232	152-225

$n_1 = 14$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**
14	147-259	152-254	160-246	166-240
15	151-269	156-264	164-256	171-249

$n_1 = 15$				
n_2	**0,005**	**0,01**	**0,025**	**0,05**
15	171-294	176-289	184-281	192-273

Fonte: MINIUM, E.W.; CLARKE, R.C.; GOLADARCI, T. Elements of Statistical Reasoning. New York: Wiley, 1999. p. 474-475.

Tabela 4. Valores críticos para *T* no teste de Wilcoxon

	Nível de significância para teste unilateral		
N	0,025	0,01	0,005
	Nível de significância para teste bilateral		
	0,05	0,02	0,01
6	0	–	–
7	2	0	–
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

Fonte: SIEGEL, S. *Nonparametric Statistics*. New York: McGraw-Hill, 1959. p. 280-281.

Tabela 5. Valores críticos para H no teste de Kruskal-Wallis

Tamanho da amostra					Tamanho da amostra				
n_1	n_2	n_3	*H*	*p*	n_1	n_2	n_3	*H*	*p*
2	1	1	2,7000	0,500	4	3	2	6,4444	0,008
2	2	1	3,6000	0,200				6,3000	0,011
								5,4444	0,046
2	2	2	4,5714	0,067				5,4000	0,051
			3,7143	0,200				4,5111	0,098
								4,4444	0,102
3	1	1	3,2000	0,300					
					4	3	3	6,7455	0,010
								6,7091	0,013
3	2	1	4,2857	0,100				5,7909	0,046
			3,8571	0,133				5,7273	0,050
			5,3572	0,029				4,7091	0,092
			4,7143	0,048				4,7000	0,101
			4,5000	0,067					
			4,4643	0,105	4	4	1	6,6667	0,010
								6,1667	0,022
3	3	1	5,1429	0,043				4,9667	0,048
			4,5714	0,100				4,8667	0,054
			4,0000	0,129				4,1667	0,082
								4,0667	0,102
3	3	2	6,2500	0,011					
			5,3611	0,032	4	4	2	7,0364	0,006
			5,1389	0,061				6,8727	0,011
			4,5556	0,100				5,4545	0,046
			4,2500	0,121				5,2364	0,052
								4,5545	0,098
								4,4455	0,103
3	3	3	7,2000	0,004					
			6,4889	0,011					
			5,6889	0,029	4	4	3	7,1439	0,010
			5,6000	0,050				7,1364	0,011
			5,0667	0,086				5,5985	0,049
			4,6222	0,100				5,5758	0,051
								4,5455	0,099
								4,4773	0,102

Continua

Tabela 5. Continuação

Tamanho da amostra					Tamanho da amostra				
n_1	n_2	n_3	*H*	*p*	n_1	n_2	n_3	*H*	*p*
4	1	1	3,5714	0,200					
4	2	1	4,8214	0,057	4	4	4	7,6538	0,008
			4,5000	0,076				7,5385	0,011
			4,0179	0,114				5,6923	0,049
4	2	2	6,0000	0,014				5,6538	0,054
			5,3333	0,033				4,6539	0,097
			5,1250	0,052				4,5001	0,104
			4,4583	0,100	5	1	1	3,8571	0,143
			4,1667	0,105	5	2	1	5,2500	0,036
4	3	1	5,8333	0,021				5,0000	0,048
			5,2083	0,050				4,4500	0,071
			5,0000	0,057				4,2000	0,095
			4,0556	0,093				4,0500	0,119
			3,8889	0,129					
5	2	2	6,5333	0,0008				4,5487	0,099
			6,1333	0,013				4,5231	0,103
			5,1600	0,034					
			5,0400	0,056	5	4	4	7,7604	0,009
			4,3733	0,090				7,7440	0,011
			4,2933	0,122				5,6571	0,049
								5,6176	0,050
5	3	1	6,4000	0,012				4,6187	0,100
			4,9600	0,048				4,5527	0,102
			4,8711	0,052					
			4,0178	0,095	5	5	1	7,3091	0,009
			3,8400	0,123				6,8364	0,011
								5,1273	0,046
5	3	2	6,9091	0,009				4,9091	0,053
			6,8218	0,010				4,1091	0,086
			5,2509	0,049				4,0364	0,105
			5,1055	0,052					
			4,6509	0,091	5	5	2	7,3385	0,010
			4,4945	0,101				7,2692	0,010
								5,3385	0,047

Continua

Tabela 5. Continuação

Tamanho da amostra					Tamanho da amostra				
n_1	n_2	n_3	*H*	*p*	n_1	n_2	n_3	*H*	*p*
5	3	3	7,0788	0,009				5,2462	0,051
			6,9818	0,011				4,6231	0,097
			5,6485	0,049				4,5077	0,100
			5,5152	0,051					
			4,5333	0,097	5	5	3	7,5780	0,010
			4,4121	0,109				7,5429	0,10
								5,7055	0,046
5	4	1	6,9545	0,008				5,6264	0,051
			6,8400	0,011				4,5451	0,100
			4,9855	0,044				4,5363	0,102
			4,8600	0,056					
			3,9873	0,098	5	5	4	7,8229	0,010
			3,9600	0,102				7,7914	0,010
								5,6657	0,049
5	4	2	7,2045	0,009				5,6429	0,050
			7,1182	0,010				4,5229	0,099
			5,2727	0,049				4,5200	0,101
			5,2682	0,050					
			4,5409	0,098					
			4,5182	0,101	5	5	5	8,0000	0,009
								7,9800	0,010
5	4	3	7,4449	0,010				5,7800	0,049
			7,3949	0,011				5,6600	0,051
			5,6564	0,049				4,5600	0,100
			5,6308	0,050				4,5000	0,102

Fonte: SIEGEL, S. Nonparametric statistics. New York: McGraw-Hill, 1959. p. 282-283.

Tabela 6. Valores críticos de q para os testes não paramétricos de comparações múltiplas

k	α:	**0,50**	**0,20**	**0,10**	**0,05**	**0,02**	**0,01**	**0,005**	**0,002**	**0,001**
2		0,674	1,282	1,645	1,960	2,327	2,576	2,807	3,091	3,291
3		1,383	1,834	2,128	2,394	2,713	2,936	3,144	3,403	3,588
4		1,732	2,128	2,394	2,639	2,936	3,144	3,342	3,588	3,765
5		1,960	2,327	2,576	2,807	3,091	3,291	3,481	3,719	3,891
6		2,128	2,475	2,713	2,936	3,209	3,403	3,588	3,820	3,988
7		2,261	2,593	2,823	3,038	3,304	3,494	3,675	3,902	4,067
8		2,369	2,690	2,914	3,124	3,384	3,570	3,748	3,972	4,134
9		2,461	2,773	2,992	3,197	3,453	3,635	3,810	4,031	4,191
10		2,540	2,845	3,059	3,261	3,512	3,692	3,865	4,083	4,241
11		2,609	2,098	3,119	3,317	3,565	3,743	3,914	4,129	4,286
12		2,671	2,965	3,172	3,368	3,613	3,789	3,957	4,171	4,326
13		2,726	3,016	3,220	3,414	3,656	3,830	3,997	4,209	4,363
14		2,777	3,062	3,264	3,456	3,695	3,868	4,034	4,244	4,397
15		2,823	3,105	3,304	3,494	3,731	3,902	4,067	4,276	4,428
16		2,866	3,144	3,342	3,529	3,765	3,935	4,098	4,305	4,456
17		2,905	3,181	3,376	3,562	3,796	3,965	4,127	4,333	4,483
18		2,942	3,215	3,409	3,593	3,825	3,993	4,154	4,359	4,508
19		2,976	3,246	3,439	3,622	3,852	4,019	4,179	4,383	4,532
20		3,008	3,276	3,467	3,649	3,878	4,044	4,203	4,406	4,554
21		3,038	3,304	3,494	3,675	3,902	4,067	4,226	4,428	4,575
22		3,067	3,331	3,519	3,699	3,925	4,089	4,247	4,448	4,595
23		3,094	3,356	3,543	3,722	3,947	4,110	4,268	4,468	4,614
24		3,120	3,380	3,566	3,744	3,968	4,130	4,287	4,486	4,632
25		3,144	3,403	3,588	3,765	3,988	4,149	4,305	4,504	4,649

Fonte: ZAR, J.H. *Biostatistical Analysis*. New Jersey: Upper Saddle River, App. 107, 1999.

Nota: $P(q) \leq \alpha \dfrac{1}{K\left(K-1\right)}$

Tabela 7 Valores críticos para $c^2{}_r$ no teste de Friedman

k = 3

***n* = 2**		***n* = 3**		***n* = 4**		***n* = 5**	
$c^2{}_r$	***p*-valor**	$c^2{}_r$	***p*-valor**	$c^2{}_r$	***p*-valor**	$c^2{}_r$	***p*-valor**
0	1,000	0,000	1,000	0,0	1,000	0,0	1,000
1	0,833	0,667	0,944	0,5	0,931	0,4	0,954
3	0,500	2,000	0,528	1,5	0,653	1,2	0,691
4	0,167	2,667	0,361	2,0	0,431	1,6	0,522
		4,667	0,194	3,5	0,273	2,8	0,367
		6,000	0,028	4,5	0,125	3,6	0,182
				6,0	0,069	4,8	0,124
				6,5	0,042	5,2	0,093
				8,0	0,0046	6,4	0,039
						7,6	0,024
						8,4	0,0085
						10,0	0,00077

n = 6		n = 7		n = 8		n = 9	
$c^2{}_r$	***p*-valor**	$c^2{}_r$	***p*-valor**	$c^2{}_r$	***p*-valor**	$c^2{}_r$	***p*-valor**
0,00	1,000	0,000	1,000	0,00	1,000	0,000	1,000
0,33	0,956	0,286	0,964	0,25	0,967	0,222	0,971
1,00	0,740	0,857	0,768	0,75	0,794	0,667	0,814
1,33	0,570	1,143	0,620	1,00	0,654	0,889	0,865
2,33	0,430	2,000	0,486	1,75	0,531	1,556	0,569
3,00	0,252	2,571	0,305	2,25	0,355	2,000	0,398
4,00	0,184	3,429	0,237	3,00	0,285	2,667	0,328
4,33	0,142	3,714	0,192	3,25	0,236	2,889	0,278
5,33	0,072	4,571	0,112	4,00	0,149	3,556	0,187
6,33	0,052	5,429	0,085	4,75	0,120	4,222	0,154
7,00	0,029	6,000	0,052	5,25	0,079	4,667	0,107
8,33	0,012	7,143	0,027	6,25	0,047	5,556	0,069
9,00	0,0081	7,714	0,021	6,75	0,038	6,000	0,057
9,33	0,0055	8,000	0,016	7,00	0,030	6,222	0,048
10,33	0,0017	8,857	0,0084	7,75	0,018	6,889	0,031
12,00	0,00013	10,286	0,0036	9,00	0,0099	8,000	0,019
		10,571	0,0027	9,25	0,0080	8,222	0,016
		11,143	0,0012	9,75	0,0048	8,667	0,010

Continua

Tabela 7. Continuação

n = 6		n = 7		n = 8		n = 9	
c^2_r	***p*-valor**	c^2_r	***p*-valor**	c^2_r	***p*-valor**	c^2_r	***p*-valor**
		12,286	0,00032	10,75	0,0024	9,556	0,0060
		14,000	0,000021	12,00	0,0011	10,667	0,0035
				12,25	0,00086	10,889	0,0029
				13,00	0,00026	11,556	0,0013
				14,25	0,000061	12,667	0,00066
				16,00	0,0000036	13,556	0,00035
						14,000	0,00020
						14,222	0,000097
						14,889	0,000054
						16,222	0,000011
						18,000	0,0000006

k = 4

n = 2		n = 3		n = 4			
c^2_r	***p*-valor**	c^2_r	***p*-valor**	c^2_r	***p*-valor**	c^2_r	***p*-valor**
0,0	1,000	0,2	1,000	0,0	1,000	5,7	0,141
0,6	0,958	0,6	0,958	0,3	0,992	6,0	0,105
1,2	0,834	1,0	0,910	0,6	0,928	6,3	0,094
1,8	0,792	1,8	0,727	0,9	0,900	6,6	0,077
2,4	0,625	2,2	0,608	1,2	0,800	6,9	0,068
3,0	0,542	2,6	0,524	1,5	0,754	7,2	0,054
3,6	0,458	3,4	0,446	1,8	0,677	7,5	0,052
4,2	0,375	3,8	0,342	2,1	0,649	7,8	0,036
4,8	0,208	4,2	0,300	2,4	0,524	8,1	0,033
5,4	0,167	5,0	0,207	2,7	0,508	8,4	0,019
6,0	0,042	5,4	0,175	3,0	0,432	8,7	0,014
		5,8	0,148	3,3	0,389	9,3	0,012
		6,6	0,075	3,6	0,355	9,6	0,0069
		7,0	0,054	3,9	0,324	9,9	0,0062
		7,4	0,033	4,5	0,242	10,2	0,0027
		8,2	0,017	4,8	0,200	10,8	0,0016
		9,0	0,0017	5,1	0,190	11,1	0,00094
				5,4	0,158	12,0	0,000072

Fonte: MINIUM, E.W.; CLARKE, R.C.; COLADARCI, T. *Elements of Statistical Reasoning*. New York: Wiley, 1999. p. 473-475.

Tabela 8. Valores críticos para r_s, o coeficiente de correlação de Spearman

n	Nível de significância para teste unilateral		Nível de significância para teste bilateral	
	0,05	**0,01**	**0,05**	**0,01**
4	0,8000	–	–	–
5	0,8000	0,9000	0,9000	–
6	0,7714	0,8857	0,8286	0,9426
7	0,6786	0,8571	0,7450	0,8929
8	0,6190	0,8095	0,7143	0,8571
9	0,5833	0,7667	0,6833	0,8167
10	0,5515	0,7333	0,6364	0,7818
11	0,5273	0,7000	0,6091	0,7545
12	0,4965	0,6713	0,5804	0,7273
13	0,4780	0,6429	0,5549	0,6978
14	0,4593	0,6220	0,5341	0,6747
15	0,4429	0,6000	0,5179	0,6536
16	0,4265	0,5824	0,5000	0,6324
17	0,4118	0,5637	0,4853	0,6152
18	0,3994	0,5480	0,4716	0,5975
19	0,3895	0,5333	0,4579	0,5825
20	0,3789	0,5203	0,4451	0,5684
21	0,3688	0,5078	0,4351	0,5545
22	0,3597	0,4963	0,4241	0,5426
23	0,3518	0,4852	0,4150	0,5306
24	0,3435	0,4748	0,4061	0,5200
25	0,3362	0,4654	0,3977	0,5100
26	0,3299	0,4564	0,3894	0,5002
27	0,3236	0,4481	0,3822	0,4915
28	0,3175	0,4401	0,3749	0,4828
29	0,3113	0,4320	0,3685	0,4744
30	0,3059	0,4251	0,3620	0,4665

Fonte: MINIUM, E.W.; CLARKE, R.C.; COLADARCI, T. *Elements of Statistical Reasoning*. New York: Wiley, 1999. p. 473-475.

Respostas dos Exercícios

CAPÍTULO 1

1.5.1. Variáveis: a) densidade óssea: contínua; b) marcas comerciais de um mesmo analgésico: variável nominal; c) comprimento de implantes: variável contínua; d) gosto (doce ou amargo): nominal; e) sexo: nominal; f) número de dentes permanentes irrompidos em uma criança: varíavel discreta; g) número de pacientes atendidos por dia em um consultório: discreta; h) altura da face: contínua; i) qualidade percebida do atendimento (bom, normal, ruim): ordinal.

1.5.2. a) O pesquisador pode conferir notas às diversas tomadas radiográficas de um mesmo corpo de prova, feitas por diferentes técnicas e/ou diferentes aparelhos. b) O pesquisador pode classificar os pacientes como já tratados ou virgens de tratamento.

1.5.3. Idade: numérico, discreto; sexo: categorizado, nominal; escolaridade: categórico, ordinal; doenças presentes: categorizado, nominal; pressão arterial: numérico, contínuo; uso de medicamentos: categórico, nominal; alérgico à aspirina: categórico, nominal; disposição em participar de um experimento: categorizado, nominal.

1.5.4. A nota 2.

1.5.5. Você pode pedir às mães para preencher as escalas, ao longo do tratamento ou após seções em branco (em que o dentista apenas apresenta o consultório ou o instrumental à criança ou só examina a criança ou mostra outra criança sendo tratada etc.). Parece mais fácil, porém, dar uma lista de palavras simples, que descrevam o que a criança expressa.

1.5.6. Parece razoável considerar que a última questão do questionário foi escrita de maneira a obter o número de não respondentes e/ou respostas inadequadas, por conta da não compreensão da questão.

1.5.7. É claro que o rato que não dormiu não o fez no tempo estipulado. O dado é censurado.

1.5.8. Sexo: categorizado, nominal; idade: numérico, discreto; peso: numérico, contínuo; estatura: numérico, contínuo; doenças: categorizado, nominal; tratamento médico: categorizado, nominal; esporte que pratica: categorizado, nominal.

1.5.9. Ordinal.

1.5.10. Falso. Talvez tenha mais dados perdidos.

CAPÍTULO 2

2.5.1. a) H_0: a nova droga produz o mesmo efeito que a tradicional contra H_1: a nova droga produz efeito diferente da tradicional; b) H_0: a dieta em questão não aumenta a longevidade contra H_1: a dieta aumenta a longevidade; c) H_0: o produto não é cancerígeno contra H_1: o produto é cancerígeno; d) H_0: a vitamina não melhora o desempenho de atletas contra H_1: a vitamina melhora o desempenho de atletas.

2.5.2. Resposta 1. Aumentar o tamanho da amostra torna o teste de hipóteses mais propenso a rejeitar a hipótese da nulidade quando é, de fato, falsa. Assim, aumenta o poder do teste. O efeito do tratamento não é afetado pelo tamanho da amostra. A probabilidade de cometer erro Tipo II fica menor, não maior, à medida que o tamanho da amostra aumenta.

2.5.3. Velocidade de leitura.

2.5.4. H_0: a moeda é bem balanceada, isto é, a probabilidade de sair cara é exatamente 0,5. Se isso for verdade, a probabilidade de saírem seis caras quando se lança a moeda é 0,015625. No entanto, se a moeda for bem balanceada e ocorrerem seis caras, você rejeita H_0: está aí o *p*-valor, 0,015625. Esse é um teste unilateral. Neste caso, H_1: a probabilidade de sair cara na moeda é maior do que 0,5. Se você resolveu por um teste unilateral, você tinha alguma razão para supor que nos lançamentos dessa moeda ocorrem mais caras. Se isso não for verdade, faça um teste bilateral. Nesse caso, H_1: a probabilidade de sair cara na moeda é diferente de 0,5 e o *p*-valor é 0,031250.

2.5.5. Para definir a população, é importante saber de onde veio a amostra (centro cirúrgico) e uma descrição dos critérios de inclusão. Por exemplo, são muitos doentes em estado grave? De que idade?

2.5.6. Rejeitar H_0 quando H_0 é falsa.

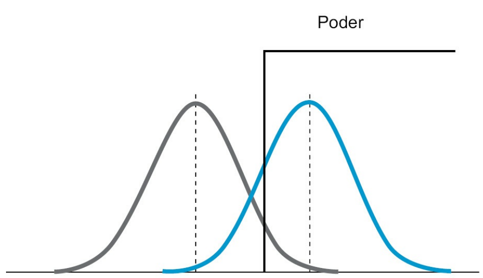
2.5.7. H_0: a droga A produz o mesmo efeito que a droga B contra H_1: a droga A produz efeito diferente da droga B.

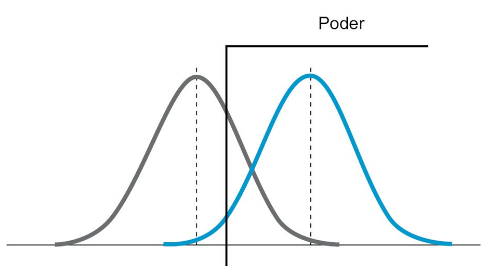
2.5.8. H_0: o nível do hormônio em estudo é o mesmo, tanto em mulheres jovens que não estão em gestação como em mulheres jovens que estão no primeiro trimestre de gestação, contra H_1: o nível do hormônio em estudo é maior em mulheres jovens que estão em gestação do que em mulheres jovens que não estão no primeiro trimestre de gestação.

Com base no resultado do teste, pode-se dizer que gestantes têm maior quantidade desse hormônio (*p*-valor = 0,001).

2.5.9. A probabilidade de o pesquisador ter obtido o resultado que obteve por acaso é $\leq$ 0,05.

2.5.10. Resposta 3, porque 1) Aumentar o tamanho da amostra torna o teste mais propenso a rejeitar a hipótese da nulidade quando é, de fato, falsa. 2) Aumentar o nível de significância reduz a região da aceitação; isso aumenta a probabilidade de rejeitar a hipótese da nulidade, aumentando assim o poder do teste. 3) Por definição, o poder do teste é igual a um menos beta, logo, o poder de um teste ficará menor à medida que o beta for maior.





CAPÍTULO 3

3.4.1. $\sum R = n \times \frac{n+1}{2} = 55$

3.4.2. a) Hipótese da nulidade: a variável tem igual distribuição nos dois grupos e a hipótese alternativa é de que de isso não acontece. b) $\sum R_1 = 25$; os valores críticos de $\sum R_1$ no nível de significância de 5% são 11 a 29. Como o valor calculado para $\sum R_1$ está entre os valores críticos, não rejeite H_0. c) As duas linhagens de ratos de laboratório não diferem quanto à habilidade de correr labirinto. d) $\sum R_2 = 20$ e

$$\sum R_1 + \sum R_2 = n \times \frac{n+1}{2} = 25 + 20 = 9 \times \frac{10}{2} = 45$$

3.4.3.

Atribuição dos postos

Grupo	**Dado**	**Posto**	**Grupo**	**Dado**	**Posto**
E	14,3	1			
E	14,6	2,5			
E	14,6	2,5			
E	14,7	4			
E	14,9	5			
			C	15,5	6,5
			C	15,5	6,5
E	15,8	8			
E	16,1	9			
E	16,4	10			
			C	16,7	11,5
			C	16,7	11,5
E	16,8	14			
			C	16,8	14
			C	16,8	14
E	17,2	16,5			
			C	17,2	16,5
			C	17,6	18
			C	17,9	19
			C	18	20
		72,5			137,5

$\sum R_1 = 72{,}5$ e $\sum R_2 = 137{,}5$; $U = 17{,}5$; $z = -2{,}45677$, p-valor = 0,014025, significante ao nível de 5%. Fazendo a correção para os empates: $z = -2{,}46419$, p-valor = 0,013737, significante no nível de 5%.

3.4.4. $\sum R_1 = 50{,}5$ e $\sum R_2 = 102{,}5$; $U = 14{,}5$; $z = -2{,}06884$, p-valor = 0,038569, significante no nível de 5%.

3.4.5.

Atribuição dos postos

Restaurante	Antes	Depois	Diferença	Posto assinalado
1	80	90	–10	–7
2	83	85	–2	–1,5
3	82	87	–5	–4,5
4	81	78	3	3
5	77	75	2	1,5
6	77	82	–5	–4,5
7	65	75	–10	–7
8	67	85	–18	–10
9	75	90	–15	–9
10	85	95	–10	–7

n = 10; T = 4,5; z = 2,34438 significante no nível de 5%, p-valor = 0,019065. O seminário foi eficaz.

3.4.6. A mediana é 55. O valor de χ^2 é 6,3427 significante no nível de 5%. O p-valor é 0,0118.

Dados em relação à mediana	Grupos		Total
	Conformistas	**Não conformistas**	
Menores ou iguais à mediana	8	2	10
Maiores que a mediana	2	7	9
Total	10	9	19

3.4.7. Como a amostra é pequena (n = 8), o certo é *não* usar a aproximação normal. Se for essa a opção, procure o valor crítico para T na Tabela 4 do Apêndice (lembre-se de que T é a menor soma de postos em valor absoluto). Para ∼ = 0,05 e amostras de tamanho n = 8, o valor crítico é 4. Como a menor soma obtida desprezando o sinal é T = 4, você pode afirmar, no nível de 5% de significância, que o ansiolítico tem efeito sobre pacientes que vão se submeter à extração de vários dentes.

3.4.8. Descartando os empates, n = 18; z = 1,65, não significante no nível de 5%.

3.4.9. ƽR_1 = 42,5 e ƽR_2 = 62,5; U = 21,5; z = – 0,323, não significante no nível de 5%.

3.4.10. ƽR_1 =23; ƽR_2 = 55; U = 2; valor crítico = 5; *significante* no nível de 5%.

CAPÍTULO 4

4.3.1.

Atribuição de postos

Grupo	Estatura	Posto	Grupo	Estatura	Posto	Grupo	Estatura	Posto
			B	1,68	1			
A	1,69	2						

A	1,7	3						
						C	1,73	4
			B	1,74	5			
A	1,75	6						
			B	1,76	7			
						C	1,77	8
A	1,78	9						
			B	1,79	10			
			B	1,8	11			
						C	1,81	12
A	1,82	13						
						C	1,84	14
						C	1,89	15
		33			34			53

n = 15; õR_1 = ÑÑŤ õR_2 = 34; õR_3 = 53; *H* = 2,54, com 2 graus de liberdade, não significante no nível de 5%.

4.3.2.

Atribuição de postos

Escore	Negro	Posto	Escore	Latino	Posto	Escore	Branco	Posto
1	N	1						
						2	B	2
3	N	3,5				3	B	3,5
			4	L	5			
			5	L	6,5			
						5	B	6,5
6	N	8						
7	N	9				8	B	10
9	N	11						
10	N	12,5						
						10	B	12,5
Escore	**Negro**	**Posto**	**Escore**	**Latino**	**Posto**	**Escore**	**Branco**	**Posto**
11	N	14						
			12	L	15,5			
						12	B	12,5
			13	L	17			
						14	B	18
			15	L	19			

16	N	20						
			17	L	21			
						18	B	22
20	N	23						
21	N	24						
			22	L	25			
						23	B	26
						25	B	27
			27	L	28			
			29	L	29			
			30	L	30			
		126			196			143

$n = 30$; $\Sigma R_1 = 126$; $\Sigma R_2 = 196$; $\Sigma R_3 = 143$; $H = 3{,}44$, corrigido para os empates, com 2 graus de liberdade, não significante no nível de 5%. O p-valor é 0,1788.

4.3.3. Mediana geral = 7,9, $\chi^2 = 9{,}60$, com 3 graus de liberdade, significante no nível de 5%; p-valor = 0,0223.

Cálculos intermediários: frequências observadas

Condição	A	B	C	D	Total
Menores ou iguais à mediana	8	9	2	6	25
Maiores que a mediana	4	3	10	6	23
Total	12	12	12	12	48

4.3.4. $\Sigma R_1 = 19{,}5$; $\Sigma R_2 = 26{,}5$; $\Sigma R_3 = 20{,}5$; $\Sigma R_4 = 33{,}5$; $\chi^2 = 7{,}25$ com 3 graus de liberdade, não significante no nível de 5%. Se puder usar computador, faça o teste com correção para os empates, o que é o mais certo. Sem a correção, você praticou os cálculos. Com a correção, $\chi^2 = 7{,}979$ com 3 graus de liberdade, significante no nível de 5%.

4.3.5 Mediana geral = 25; $\chi^2 = 5{,}57$, com 2 graus de liberdade, não significante no nível de 5%.

4.3.6 $\Sigma R_1 = 21$; $\Sigma R_2 = 38$; $\Sigma R_3 = 61$; υ = 8,06 com 2 graus de liberdade, significante no nível de 5%. O p-valor é 0,00178. Para o teste de Dunn

$$s = \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{\frac{15 \times 16}{12}\left(\frac{1}{5} + \frac{1}{5}\right)} = \sqrt{20 \times \frac{2}{5}} = 2{,}828427$$

Tabela resumo dos resultados do teste de Dunn

Comparação	Diferença dos postos médios	Q	Q crítico	Conclusão
A versus B	–3,4	–1,202081	2,394	Não rejeita H_0
A versus C	–8,0	–2,828427	2,394	Rejeita H_0
B versus C	–4,6	–1,626345	2,394	Não rejeita H_0

4.3.7.

Atribuição dos postos

Grupo I	Grupo II	Grupo III
4	16	10
7	18	20
3	15	2
8	11	22
14	17	21
9	13	19
6		
1		
5		
12		
$SR_1 = 69$	$SR_2 = 90$	$SR_3 = 94$

H = 9,232 com 2 graus de liberdade, significante no nível de 1%. Postos médios: grupo I: 6,90; grupo II: 15,00; grupo III: 15,67.

Tabela resumo dos resultados do teste de Dunn

Comparação	Diferença dos postos médios	*Q*	*Q* crítico	Conclusão
I versus II	8,1	2,42	2,394	Rejeita H_0
I versus III	8,77	2,42	2,394	Rejeita H_0
II versus III	0,67	2,16	2,394	Não rejeita H_0

4.3.8. Mediana geral = 33,0; χ^2 = 25,53, com 3 graus de liberdade, significante no nível de 5%.

Condição	Comunidade				Total
	Muito isolada	Isolada	Rural	Favela	
Menor ou igual à mediana	21	13	8	5	47
Maior do que a mediana	2	10	15	18	45

4.3.9. ƽR_1 = 19,5; ƽR_2 = 9,0; ƽR_3 = 25,5; χ^2 = 15,94 com 2 graus de liberdade, significante no nível de 5%.

Tabela resumo para o teste de comparações múltiplas para o exemplo 6.6, ~ = 0,10

Comparação	*Ri – Rj*	Valor crítico	Conclusão
A versus B	|19,5 – 9| = 10,5	8,697	Rejeita H_0
A versus C	|19,5 – 25,5| = 6	8,697	Não rejeita H_0
B versus C	|9 – 25,5| = 16,5	8,697	Rejeita H_0

Conclui-se que A e B dão resultados diferentes, assim como B e C.

4.3.10. ƽR_1 =29; ƽR_2 =21; ƽR_3 = 10; χ^2 = 18,20 com 2 graus de liberdade, significante no nível de 5%.

CAPÍTULO 5

5.5.1 O estudo é transversal. A hipótese da nulidade é a de que a maneira de descartar o lixo não depende do sexo. A hipótese alternativa é a de que a maneira de descartar o lixo depende do sexo.

O valor de χ^2 sem correção de continuidade é 2,000, *p*-valor 0,1573. O valor de χ^2 com correção de continuidade é 1,389, *p*-valor 0,2386. Com a correção, diminui a estatística (de 2,000 para 1,389).

5.5.2 O estudo é transversal. A hipótese da nulidade é a de que o medo de viajar de avião não está associado a partida ou chegada. O valor de χ^2 sem correção de continuidade é 8,30, *p*-valor 0,0040. Portanto, existe associação. Passageiros revelam mais medo ao chegar.

Proporções dos passageiros entrevistados segundo a situação de chegada ou partida e o medo de voar

Passageiros	**Medo**		**Total**
	Sim	**Não**	
Chegavam	0,162	0,438	0,6
Partiam	0,064	0,336	0,4
Total	0,226	0,774	1

5.5.3 Os dados são de um estudo transversal. A hipótese da nulidade é de que a prevalência de sobrepeso não está associada à geração. Então foi construída a tabela dada em seguida. O valor de χ^2 sem correção de continuidade é 24,21, *p*-valor = 0,0000. A segunda geração apresentou prevalências elevadas de sobrepeso, significativamente maiores do que a primeira.

Distribuição percentual dos nipo-brasileiros com sobrepeso,segundo a geração (Bauru SP, 2000)

Geração	**Sobrepeso**		**Total**
	Sim	**Não**	
Primeira	0,062	0,132	0,193
Segunda	0,395	0,412	0,807
Total	0,456	0,544	1

5.5.4 Os dados são de um ensaio clínico. A hipótese da nulidade é a de que a prevalência de dor é a mesma nos dois grupos. O teste de proporções bilateral com correção de continuidade fornece o valor *z* = 2,545, significante no nível de 5%. O *p*-valor é 0,0109. A proporção dos participantes da pesquisa sem dor no grupo controle é 0,571 e no grupo experimental é 0,118. Portanto, o uso de betametasona diminui a prevalência de dor no pós-operatório de pacientes submetidos a tratamento endodôntico.

5.5.5 Os dados são de um ensaio clínico. A hipótese da nulidade é a de que a taxa de sobrevivência é a mesma nos dois grupos. Amostra pequena, teste exato de Fisher.

Valores observados

Grupo	Vivos 5 anos depois		Total
	Sim	Não	
Tratado	4	1	5
Controle	2	3	5
Total	6	4	10

Valores mais extremos do que os observados

Grupo	Vivos 5 anos depois		Total
	Sim	Não	
Tratado	5	0	5
Controle	1	4	5
Total	6	4	10

A probabilidade de os valores observados ocorrerem é de 0,2381. A probabilidade de ocorrerem valores mais extremos do que os observados é de 0,0238. Então *p*-valor = 0,2619. Os dados não trazem evidência de que o tratamento ajudou.

5.5.6 Os dados são de um estudo transversal. A hipótese em teste é a de que a prevalência de gengivite não está associada ao hábito de fumar. O valor do χ^2 para o teste de qui-quadrado sem correção de continuidade é 2,747, *p*-valor 0,0975, e com correção de continuidade é 2,101, *p*-valor 0,1472; os dois resultados são não significantes ($p < 0,05$).

5.5.7 Para testar H_0, de que as respostas são consistentes, é preciso reconstruir a tabela, de maneira que *a* e *d* sejam mudanças de respostas.

Tinham dor antes	Tinham dor depois		Total
	Não	Sim	
Sim	785	380	1.165
Não	215	75	290
Total	1.000	455	1.455

$$\chi^2 = \frac{(|a-d|-1)^2}{a+d}$$

$$\chi^2 = \frac{(|785-75|-1)^2}{785+75} = 584{,}513$$

O resultado é extremamente significante ($p < 0,0001$). Então, o tratamento teve efeito: 91,279% dos que tinham dor passaram a não ter, enquanto 8,728% dos que não tinham dor passaram a ter.

5.5.8 O estudo é transversal. A hipótese da nulidade é a de que o distúrbio da articulação temporomandibular (DTM) não depende de sexo.

Distribuição de 604 crianças e adolescentes de acordo com sexo e distúrbio da articulação temporomandibular (DTM)(Proporção entre parênteses)

Sexo	DTM	Total

	Sim	Não	
Masculino	118 (0,195)	146 (0,242)	264 (0,437)
Feminino	177 (0,293)	163 (0,270)	340 (0,563)
Total	295 (0,488)	309 (0,512)	604 (1,000)

O valor de qui-quadrado sem correção de continuidade é 3,220, *p*-valor = 0,0726. O valor de qui-quadrado com correção de continuidade é 2,940, *p*-valor = 0,0867. Não se rejeita a hipótese de que sexo e distúrbio da articulação temporomandibular (DTM) sejam independentes.

5.5.9

Distribuição das mulheres dos dois grupos segundo o uso ou não de aspirina

Grupo	Aspirina		Total	Proporção de usuárias
	Sim	Não		
Tratado	1.623	6.883	8.506	0,191
Controle	1.631	6.471	8.102	0,201
Total	3.254	13.354	16.608	0,196

A proporção de usuárias de aspirina é de 19,1% no grupo tratado e de 20,1% no grupo controle. Não há diferença estatística quanto a essa característica (p = 0,0883).

5.5.10 O estudo é transversal. A hipótese da nulidade é a de que a etiologia do trauma de face não depende de sexo. O valor de qui-quadrado com correção de continuidade é 2,27, *p*-valor = 0,1323. Não se rejeita a hipótese de que sexo e etiologia do trauma de face sejam independentes.

CAPÍTULO 6

6.6.1. O valor calculado de χ^2 é 13,62, com 3 graus de liberdade. A tabela de χ^2 dá, para o nível de significância de 1% e com 3 graus de liberdade, o valor 11,34. Como o valor calculado é maior do que o da tabela, conclui-se que o tipo sanguíneo está associado à origem.

6.6.2. O valor calculado de χ^2 é 16,86, com 7 graus de liberdade. A tabela de χ^2 dá, para o nível de significância de 5% e com 7 graus de liberdade, o valor 14,07. Como o valor calculado é maior do que o valor crítico, pode-se concluir que a confirmação da denúncia está associada ao perfil do denunciante.

Percentuais de denúncias confirmadas e não confirmadas segundo o perfil do notificante

Perfil do notificante	Denúncia	
	Não confirmada	Confirmada
Familiares	24,00%	8,50%
Amigos e vizinhos	23,90%	3,70%
Anônimo	14,30%	2,40%
Desconhecido	9,40%	1,00%
Profissionais	3,70%	0,50%

A própria criança	3,40%	0,30%
Outros	4,10%	0,70%
Total	82,80%	17,20%

6.6.3. Para analisar os dados aplicando o teste χ^2, é preciso reunir disfunção moderada e disfunção severa em uma só categoria. No entanto, seria melhor – o que nem sempre o tempo e o orçamento permitem – aumentar o tamanho da amostra.

6.6.4. Testar a hipótese de independência.

6.6.5. O valor calculado de χ^2 é 4,16, com 2 graus de liberdade. O valor crítico de χ^2 com 2 graus de liberdade e no nível de significância de 5% é 5,99. Não se detectou associação entre a resposta do paciente e a combinação de drogas. A partição da tabela foi, então, feita como mostrado em seguida. O valor do χ^2, com 1 grau de liberdade, é 3,85, significante no nível de 5%. O *p*-valor é 0,0498. A combinação com candesartan deu, proporcionalmente, mais respostas positivas.

Pacientes com resposta positiva ou negativa segundo a combinação de drogas

Droga	**Resposta do paciente**		**Total**	**Percentual de respostas positivas**
	Positiva	**Negativa**		
Candesartan	95	44	139	68,30%
Losartan	75	57	132	56,80%
Total	170	101	271	62,70%

6.6.6. O estudo é transversal. A hipótese da nulidade é de que a resposta "Não" independe do tipo de profissional. O valor de χ^2 é 12,1307, significante no nível de 5%. O *p*-valor é 0,016405. A hipótese da nulidade foi, portanto, rejeitada.

Cálculos auxiliares para obtenção do valor de χ^2

Tipo de profissional	***O***	***E***	**(*O*-*E*)**	**(*O*-*E*)^2**	**(*O*-*E*)^2/*E***
	Respostas "Sim"				
Enfermeiro	264	255	9	81	0,31765
Auxiliar de enfermagem	254	255	–1	1	0,00392
Parteira	258	255	3	9	0,03529
Técnico	237	255	–18	324	1,27059
Outros	262	255	7	49	0,19216
	Respostas "Não"				
Enfermeiro	36	45	–9	81	1,80000
Auxiliar de enfermagem	46	45	1	1	0,02222
Parteira	42	45	–3	9	0,20000
Técnico	63	45	18	324	7,20000
Outros	38	45	–7	49	1,08889
Total	1.500	1.500	0		12,13072

Para fazer o procedimento de Marascuilo, é preciso calcular:

$$d_{ij} = \sqrt{\chi^2_{(s-1)}}\sqrt{\frac{p_i(1-p_i)}{n_i} + \frac{p_j(1-p_j)}{n_j}}$$

Adotando o nível de 5% de significância, o valor crítico de χ^2 com 4 graus de liberdade é 9,488 e a raiz quadrada desse valor é 3,080.

Uma diferença é estatisticamente significante se seu valor excede o valor da amplitude crítica. Neste exemplo, embora o teste de χ^2 tenha rejeitado a hipótese de que a resposta "Não" independe do tipo de profissional, o procedimento de Marascuilo não detectou onde está a diferença. O contato entre enfermeiro e técnico se aproxima, porém, da amplitude crítica, indicando que uma amostra maior talvez mostrasse a diferença entre esses dois tipos de profissionais.

Tabela-resumo para o procedimento de Marascuilo

Comparação	**Contraste**	**Amplitude crítica**	**Conclusão**
E - A	0,033	0,086	Não
E - P	0,020	0,085	Não
E - T	0,090	0,093	Não
E - O	0,007	0,083	Não
A - P	0,013	0,089	Não
A - T	0,057	0,097	Não
A - O	0,026	0,087	Não
P - T	0,070	0,095	Não
P - O	0,013	0,086	Não
T - O	0,083	0,094	Não

6.6.7. O estudo é transversal. A hipótese da nulidade é a de que a idade em que se começa a usar lentes de contato não depende do sexo.

Frequências esperadas de usuários de lentes de contato classificados de acordo com o sexo e a idade em que começaram a usar esse tipo de lente

Sexo	**Faixa de idade, em anos**			
	Menos de 15	**De 15 a 19**	**20 e mais**	**Total**
Homens	3,4831	45,6292	12,8876	62
Mulheres	6,5169	85,3708	24,1124	116
Total	10	131	37	178

Valores calculados para $(O - E)^2/E$

Sexo	**Faixa de idade**			**Total**
	Menos de 15	**De 15 a 19**	**20 e mais**	
Homens	0,631533164	1,275605997	6,4430023	8,350141
Mulheres	0,337543588	0,681789412	3,4436736	4,463007
Total	0,969076752	1,95739541	9,8866759	12,81315

O valor de χ^2 (12,81315) é significante a 5%. Rejeita-se H_0.

6.6.8. O estudo é transversal. A hipótese da nulidade é a de que a resposta independe do grupo. O valor de χ^2 é 7,4063 com (3-1) (4-1) = 6 graus de liberdade, não significante no nível de 5%. O *p*-valor é 0,2849. Veja as proporções de respostas "Sim".

O médico cirurgião HIV-positivo pode continuar exercendo sua profissão?

Respondentes		**Sim**	**Não**	**Não declarou**	**Total**	**Proporção de "Sim"**
Estudantes	Básico	38	17	5	60	63,30%
	Clínico	49	8	3	60	81,70%
Médicos	Residentes	16	2	2	20	80,00%
	Pós-residentes	23	6	1	30	76,70%
Total		126	33	11	170	74,10%

Cálculos intermediários para aplicar o teste de χ^2

Grupo	***O***	***E***	**(*O*-*E*)**	**(*O*-*E*)^2**	**(*O*-*E*)^2/*E***
			Resposta "Sim"		
Estudante básico	38	44,47059	–6,47059	41,8685	0,9415
Estudante clínico	49	44,47059	4,529412	20,5156	0,4613
Médico residente	16	14,82353	1,176471	1,3841	0,0934
Pós-residente	23	22,23529	0,764706	0,5848	0,0263
			Resposta "Não"		
Estudante básico	17	11,64706	5,352941	28,6540	2,4602
Estudante clínico	8	11,64706	–3,64706	13,3010	1,1420
Médico residente	2	3,882353	–1,88235	3,5433	0,9127
Pós-residente	6	5,823529	0,176471	0,0311	0,0053
			Sem declaração		
Estudante básico	5	3,882353	1,117647	1,24914	0,32175
Estudante clínico	3	3,882353	–0,88235	0,77855	0,20054
Médico residente	2	1,294118	0,705882	0,49827	0,38503
Pós-residente	1	1,941176	–0,94118	0,88581	0,45633
Total	170	170	0		7,40633

Há frequências esperadas menores do que 5. Então, o teste não deve ser aplicado. No entanto, o resultado é não significante. Como questionários sem resposta à pergunta foram poucos, isto é, 11 em 170, o que significa 6,5% da amostra, é razoável desprezar esses dados e fazer o teste com respostas "Sim" e "Não".

Cálculos intermediários para aplicar o teste de χ^2

Grupo	***O***	***E***	**(*O*-*E*)**	**(*O*-*E*)^2**	**(*O*-*E*)^2/*E***
			Resposta "Sim"		
Estudante básico	38	43,58	-5,58	31,1364	0,7145
Estudante clínico	49	45,17	3,83	14,6689	0,3247

Médico residente	16	14,26	1,74	3,0276	0,2123
Pós-residente	23	22,98	0,02	0,0004	0,0000
			Resposta "Não"		
Estudante básico	17	11,42	5,58	31,1364	2,7265
Estudante clínico	8	11,83	-3,83	14,6689	1,2400
Médico residente	2	3,74	-1,74	3,0276	0,8095
Pós-residente	6	6,02	-0,02	0,0004	0,0001
Total	159	159	0		6,0276

Há frequência esperada menor do que 5. Então, o teste não deve ser aplicado. No entanto, o resultado é não significante. Compare respostas dos estudantes com respostas de médicos em uma tabela 2 $^{3}_{2}$ 2. Com correção de continuidade, o resultado é qui-quadrado igual a 0,289, não significante.

6.6.9. a) O valor calculado de χ^2 é 16,024, com 3 graus de liberdade. O valor crítico de χ^2, para o nível de significância de 1% e com 3 graus de liberdade, é 11,34. Como o valor calculado é maior do que o valor crítico, conclui-se que existe associação entre o tamanho do carcinoma de células renais e o tipo de detecção. É fácil ver na tabela a seguir que o percentual de tumores detectados por sintomas aumenta em função do tamanho.

Carcinoma de células renais segundo o tamanho do tumor, em centímetros, e o tipo de detecção

Tipo de detecção	**Tamanho do tumor (cm)**			
	De 0,5 a 4	**De 4 a 7**	**De 7 a 10**	**Maior que 10**
Incidental	30	21	5	3
Sintomático	11	22	14	9
Total	41	43	19	12
Percentual (sintoma)	26,8	51,2	73,7	75

b) O valor de χ^2_{tend} com 1 grau de liberdade é 2,797. O valor crítico de χ^2 no nível de significância de 5% e com 1 grau de liberdade é 3,84.

Como o valor calculado é menor do que o valor crítico, os dados não foram suficientes para mostrar *tendência estatística* de a detecção por sintomas aumentar à medida que aumenta o tamanho do carcinoma de células renais.

6.6.10. O valor calculado de χ^2 é 8,86, com 1 grau de liberdade. O valor crítico de χ^2 no nível de significância de 5% e com 1 grau de liberdade é 3,84. O valor calculado é maior do que o valor crítico de χ^2, sugerindo o efeito da amidalectomia no desenvolvimento de mononucleose.

CAPÍTULO 7

7.4.1. χ^2 = 3,0312, não significante no nível de 5%; coeficiente **$** = 0,266. Associação pequena.

7.4.2. χ^2 = 23,36 sem correção, significante no nível de 1%; *p*-valor = 0,0000; χ^2 = 22,93 com correção de Yates, significante no nível de 1%; *p*-valor = 0,0000; coeficiente **$** = 0,1066. Então, embora o valor do teste possa parecer impressionante, a associação é pequena: revelou-se significante devido ao tamanho da amostra. Note que o percentual de meninos com hábitos orais nocivos na amostra é 22,3% e o percentual de meninas é 26,5%, valores bastante próximos.

7.4.3. χ^2 = 6,51 sem correção, significante no nível de 5%; *p*-valor = 0,0107; χ^2 = 5,81 com correção de Yates, significante no nível de 5%; *p*-valor = 0,0160; coeficiente gama = 0,3496. Existe associação positiva entre as escolaridades das mulheres e seu conhecimento sobre o exame Papanicolaou.

7.4.4. χ^2 = 8,81 sem correção, significante no nível de 5%; *p*-valor = 0,0030; χ^2 = 5,45 com correção de Yates, significante no nível de 5%; *p*-valor = 0,0196; risco de óbito neonatal no caso de gestantes diabéticas = 12,5%; risco de óbito neonatal no caso de gestantes não diabéticas = 2,5%; risco relativo = 5,06, ou seja, o risco de óbito neonatal é 5 vezes maior no caso de gestantes diabéticas.

7.4.5. As medidas de associação são: $ T̀ íG̀Ḿm̥Ṕ; coeficiente de contingência *P* = 0,123; *V* de Cramer = 0,123.

7.4.6. r_s = – 0,19048, não significante no nível de 5%; *p*-valor = 0,6514.

7.4.7. a) Os dois juízes que têm opiniões mais parecidas são A e B. b) Os dois juízes que têm opiniões bem diferentes são B e C.

Coeficientes de correlação segundo os juízes

Juízes	**Coeficiente de correlação**	***p*-valor**
A e B	0,6121	0,06
A e C	–0,0545	0,881
B e C	–0,1758	0,6272

7.4.8. χ^2 = 9,83 sem correção, significante no nível de 1%; *p*-valor = 0,0017; χ^2 = 9,11 com correção de Yates, significante no nível de 1%; *p*-valor = 0,0026; coeficiente $ = 0,1457. A associação é pequena: o teste revelou-se significante devido ao tamanho da amostra.

7.4.9. χ^2 = 5,34, não significante no nível de 5%; *p*-valor = 0,2544; coeficiente de Cramér = 0,0943.

7.4.10. χ^2 = 16,86, significante no nível de 5%; *p*-valor = 0,00981; coeficiente de Cramér = 0,16953. Existe associação entre a denúncia confirmada e o perfil do denunciante. Veja que denúncias de familiares são mais confiáveis.

Denúncias confirmadas e não confirmadas segundo o perfil do notificante

Perfil do notificante	**Denúncia**		**Total**	**Percentual**
	Não confirmada	**Confirmada**		
Familiares	141	50	191	26,20%
Amigos e vizinhos	140	22	162	13,60%
Anônimo	84	14	98	14,30%
Desconhecido	55	6	61	9,80%
Profissionais	22	3	25	12,00%
A própria criança	20	2	22	9,10%
Outros	24	4	28	14,30%
Total	486	101	587	17,20%

CAPÍTULO 8

8.5.1. A sensibilidade (S), a especificidade (E) e a acurácia (A) do teste são, respectivamente:

$$S = \frac{120}{176} = 0,682$$

$$E = \frac{250}{274} = 0,912$$

$$A = \frac{120 + 250}{450} = 0,822$$

8.5.2. O valor preditivo de teste positivo (VPP) e o valor preditivo de teste negativo (VPN) são, respectivamente:

$$VPP = \frac{120}{144} = 0,833$$

$$VPN = \frac{250}{306} = 0,817$$

8.5.3.

$$NNT = \frac{1}{p_t - p_c}$$

p_t a proporção de sucessos no grupo tratado = 99/100 = 0,99

p_c a proporção de sucessos no grupo controle = 98/100 = 0,98

$$NNT = \frac{1}{p_t - p_c} = \frac{1}{0,99 - 0,98} = 100$$

8.5.4. A sensibilidade é 66,7%, a especificidade é 89,7%, o valor preditivo positivo é 51,7%, o valor preditivo negativo é 94,2% e a precisão é 86,5%. Os resultados do PSA total são mais seguros do que os do exame físico prostático.

8.5.5. Sensibilidade = 71/176 = 0,40; especificidade = 270/274 = 0,98; valor preditivo positivo = 71/75 = 0,95; valor preditivo negativo = 270/375 = 0,72; razão de verossimilhança = 20.

8.5.6. Vamos pressupor arbitrariamente que a população tenha 10.000 pessoas. Só estamos tratando razões, então o tamanho da população é arbitrado. Como a prevalência da doença é de 5%, então 0,05 × 10.000 = 500 pessoas têm a doença. Logo, 10.000 – 500 = 9.500 não têm a doença. O número de pessoas doentes que tiveram resultado positivo no teste é 0,99 × 500 = 495. O número de pessoas sem a doença que tiveram resultado negativo no teste é 0,92 × 9.500 = 8.740. O valor preditivo do exame positivo é 39,4% e o valor preditivo do exame negativo é 99,9%.

8.5.7.

Cálculos intermediários

Resultado do exame	Doença		Total	Valor preditivo
	Presente	Ausente		
Positivo	495	760	1.255	$VPP = \frac{495}{1255} = 0,394$

Negativo	5	8.740	8.745	$VPN = \frac{8740}{8745} = 0{,}999$
Total	500	9.500	10.000	
	$S = \frac{495}{500} = 0{,}99$	$E = \frac{8740}{9500} = 0{,}92$		

8.5.8. I_O = 0,450, I_E =0,455; *kappa* = – 0,00917. Note que o índice *kappa* é negativo, indicando muita discordância entre os professores na avaliação dos alunos.

Classificação dos alunos por dois professores

Professor 1	**Professor 2**			**Total**
	A	**B**	**C**	
A	8	6	0	14
B	2	1	3	6
C	0	0	0	0
Total	10	7	3	20

8.5.9. Somente quando os examinadores concordam em todas as avaliações.

8.5.10. Somente quando, feito o exame para diagnóstico, todas as pessoas doentes têm resultado positivo e todas as pessoas sem a doença em estudo têm resultado negativo.

Glossário

Acaso: (estatística) 1. termo usado para descrever os resultados de um processo estocástico, isto é, um processo no qual a probabilidade de ocorrer qualquer evento é conhecida ou pode ser determinada. 2. diz-se do resultado da soma de um complexo de numerosas causas cujas atuações individuais desconhecemos. 3. ao acaso: não significa, em estatística, a esmo, sem reflexão, inadvertidamente, mas o contrário: significa processo construído para que cada resultado possível esteja associado a uma probabilidade conhecida.

Acaso: (geral) 1. acontecimento incerto ou imprevisível; casualidade, eventualidade. 2. fortuito. 3. destino, fortuna, sorte. 2. ao acaso: a esmo, sem reflexão, inadvertidamente.

Aleatório: (estatística) 1. que acontece ao acaso, ou seja, diz-se da variável que assume valores segundo uma determinada lei de probabilidades. Por exemplo, os resultados de um jogo de dados são aleatórios. 2. quando é determinado por um complexo de numerosas causas somadas, mas cujas atuações individuais desconhecemos. Por exemplo, erro aleatório. 3. diz-se do processo construído para que cada resultado possível esteja associado a uma probabilidade conhecida. Por exemplo, em um experimento, os tratamentos são designados aos pacientes por processo aleatório.

Alocação: processo de alocar ou designar um tratamento a uma unidade experimental.

Amostra: qualquer conjunto cujas características ou propriedades são estudadas com o objetivo de estendê-las a outro conjunto, do qual o primeiro conjunto é considerado parte.

Análise de variância: técnica estatística que subdivide a variabilidade total de um conjunto de dados em seus componentes. Pode estabelecer, por exemplo, que as médias de vários grupos são estatisticamente diferentes.

Apuração de dados: processo de tomar os dados brutos, registrados em fichas clínicas físicas ou digitalizadas ou cadernos de laboratório, e organizá-los de forma satisfatória para posterior tabulação e análise.

Banco de dados: coleção ou arquivo de dados organizados de maneira específica e só acessado por pessoal com a necessária competência, para propósito definido.

Cálculo do tamanho da amostra: cálculo matemático, feito geralmente quando o ensaio é planejado, que estabelece o número de pacientes que deve ser recrutado, a um dado nível de significância e um dado poder de teste.

Casual: o mesmo que aleatório.

Casualização: procedimento adotado nos ensaios clínicos casualizados; consiste em designar, por processo aleatório – e nunca por escolha –, tratamentos pré-escolhidos aos participantes

da pesquisa. Veja randomização.

Casuística: registro pormenorizado de casos clínicos das doenças.

Comparação de tratamentos: qualquer comparação que envolva dois ou mais tratamentos ou grupos.

Comparações múltiplas: referem-se ao fato de que dois ou mais tratamentos devem ser comparados, sempre em relação à mesma variável, em determinado momento do ensaio (em geral, no fim).

Controle negativo: tratamento sem qualquer efeito farmacológico ou fisiológico, isto é, placebo ou pseudoprocedimento. Veja controle positivo.

Controle positivo: normalmente o tratamento padrão, mas sempre um tratamento que envolve o uso de uma substância farmacologicamente ativa. Veja controle negativo.

Dado censurado: é o dado cujo valor está além de um limite conhecido, embora não se saiba o valor exato do dado.

Dado numérico: é expresso por números. Veja também dado quantitativo.

Dados binários: referem-se a uma variável binária, que só assume um de dois valores possíveis, 0 ou 1.

Dados brutos: 1. medidas e observações registradas em fichas clínicas ou cadernos de laboratório, mas ainda não organizadas para interpretação. 2. listagens de dados obtidos em computador, mas na forma como foram coletados, antes de edição, resumo e análise.

Dados categorizados: são aqueles distribuídos em categorias mutuamente exclusivas. Veja também dados qualitativos.

Dados contínuos: referem-se a uma variável contínua. São obtidos por medição.

Dados discretos: referem-se a uma variável discreta. Surgem de processos de contagem.

Dados qualitativos: são aqueles distribuídos em categorias mutuamente exclusivas. Veja também dados categorizados.

Dados quantitativos: são expressos por números. Veja também dado numérico.

Dados: informação efetiva na forma de medidas, observações ou estatísticas, usada como base para argumentação.

Delineamento: a parte do ensaio que especifica os tratamentos que serão avaliados, as unidades

experimentais, a variável em análise e o modo como os tratamentos serão designados às unidades experimentais. Veja também desenho.

Desenho: o mesmo que delineamento. É usado porque tem sonoridade similar a *design*, o termo da língua inglesa que traduz. No entanto, o termo delineamento é mais adequado.

Discrepante (*outlier*): valor, leitura ou medida fora dos limites esperados e, por isso, colocado em dúvida ou considerado erro.

Distribuição casual dos tratamentos: processo de designar os tratamentos aos pacientes ao acaso usando, por exemplo, uma tabela de números aleatórios. Esse procedimento só é adotado em ensaios clínicos casualizados. Veja distribuição randômica dos tratamentos.

Distribuição randômica dos tratamentos: o mesmo que distribuição casual dos tratamentos.

Edição de dados: processo de revisar dados com a finalidade de detectar deficiências ou erros no modo como eles foram registrados ou colecionados.

Efeito do tratamento: em ensaios clínicos, a diferença entre os resultados observados no grupo experimental e no grupo controle submetido a placebo.

Efeito placebo: efeito produzido por placebo.

Ensaio clínico controlado e casualizado – *randomized controlled clinical trial* (RCCT): ensaio clínico que envolve pelo menos um tratamento em teste e um tratamento controle, com recrutamento e seguimento simultâneos de todos os grupos, em que os tratamentos são designados aos pacientes por processo aleatório, de tal maneira que nem os pacientes nem os responsáveis pela seleção e pelo tratamento desses pacientes possam influenciar a alocação de tratamentos e no qual as alocações permanecem desconhecidas dos pacientes e do pessoal clínico até o final. A alocação é conhecida dos pacientes e dos clínicos apenas por códigos, de preferência numéricos.

Ensaio clínico: atividade de pesquisa em que um ou mais seres humanos são alocados aleatoriamente a uma ou mais intervenções (que podem incluir placebo ou outro tipo de controle) para avaliar o efeito dessas intervenções em resultados biomédicos ou comportamentais relacionados à saúde. Na maioria dos casos, a unidade experimental é o homem, mas pode ser um animal experimental. Veja unidade experimental.

Ensaio: qualquer ação experimental feita com a finalidade de obter dados para julgamento ou conclusão. O mesmo que experimento.

Entrada de dados: processo de teclar os dados para armazenamento eletrônico.

Erro tipo I: consiste em rejeitar a hipótese da nulidade quando ela é verdadeira.

Erro tipo II: consiste em aceitar a hipótese da nulidade quando ela é falsa.

Estatística de teste: fórmula ou algoritmo usado para um teste de significância; o valor numérico calculado por essa fórmula ou esse algoritmo, para um teste específico de significância, usando um conjunto de dados.

Estudo: termo genérico, usado para indicar uma grande variedade de atividades de pesquisas que envolvem coleção, análise e interpretação de dados. Também usado como um sinônimo para ensaio clínico.

Estudo comparativo: estudo que envolve dois ou mais grupos de pacientes para comparar e julgar a influência de algum fator, procedimento ou alguma condição ou característica, presentes ou aplicados a um dos grupos, mas não ao outro. Sinônimo de ensaio clínico se o estudo exige a comparação de tratamentos diferentes que envolvam pacientes tratados no mesmo período de tempo.

Estudo coorte: estudo que envolve a identificação de um grande número de pessoas (coorte), algumas expostas a um fator causal suspeito, outras não expostas a esse fator. Essas pessoas são acompanhadas durante um período de tempo relativamente longo para verificar se ocorreu ou não um resultado ou uma condição de interesse, normalmente doença ou morte. Depois se comparam as proporções de ocorrências (doença ou morte) nos dois grupos, isto é, nas pessoas expostas ao fator causal suspeito e nas não expostas. Também chamado estudo prospectivo.

Estudo de caso-controle: estudo que envolve a identificação de pessoas com uma doença ou condição de interesse (casos) e de um grupo comparável de pessoas sem a doença ou condição de interesse (controles). Casos e controles são comparados com respeito a algum atributo existente, passado ou de exposição que se acredita estar relacionado com a doença ou condição. Também chamado estudo retrospectivo.

Estudo piloto: estudo preliminar projetado para indicar se um estudo maior é viável. Também usado para estabelecer o tamanho da amostra.

Estudo prospectivo: estudo no qual pessoas com uma característica ou um atributo específicos são identificadas e observadas por um período de tempo para verificar se ocorreu ou não um resultado ou uma condição de interesse, normalmente doença ou morte.

Estudo retrospectivo: estudo no qual pessoas com uma característica ou uma doença são identificadas e questionadas para saber se foram ou não expostas a determinado fator.

Experimento: trabalho científico que se destina a verificar um fenômeno físico; ensaio,

tentativa.

Experimentos multicêntricos: experimentos conduzidos em dois ou mais centros, sempre com um protocolo comum, mas com uma administração central e um centro único para receber e processar os dados.

Fator de risco: exposição ambiental, característica pessoal ou evento que afete a probabilidade de contrair determinada doença ou experimentar mudança no estado de saúde. Uma análise dos fatores de risco normalmente implica algum tipo de análise estatística para apontar ou identificar fatores de risco para determinada doença ou condição.

***Follow-up*:** seguimento do paciente.

Grupo controle: em um ensaio clínico, grupo de pacientes designados para o tratamento controle. Serve como base de comparação para o grupo que recebe o tratamento em teste.

Grupo experimental: em um ensaio clínico, é o grupo de pacientes designados ao tratamento em teste. É contrastado com o grupo controle para chegar a uma conclusão sobre um fator, uma condição ou um tratamento.

Grupo tratado: o mesmo que grupo experimental.

Hipótese alternativa: alternativa para a hipótese da nulidade, que postula haver diferença entre as populações ou os grupos em comparação, com relação ao fator, à característica ou à condição de interesse. Veja hipótese da nulidade.

Hipótese da nulidade: hipótese que postula não haver diferença entre as populações ou os grupos em comparação com relação ao fator, à característica ou à condição de interesse. Veja hipótese nula.

Hipótese nula: tradução corrente, mas equivocada, de *null hypothesis*, uma vez que não é a hipótese que tem a qualidade de nula, mas sim o que ela postula (diferença nula). Veja hipótese da nulidade.

Intervenção: termo relacionado com a definição de ensaio clínico; é uma manipulação do participante ou do ambiente do participante com a finalidade de modificar um ou mais processos biomédicos ou comportamentais relacionados à saúde e/ou a desfechos. Exemplos incluem: drogas; dispositivos biológicos; procedimentos (p. ex., técnicas cirúrgicas); entrevistas face a face; estratégias para mudar o comportamento relacionado à saúde (p. ex., dieta, terapia cognitiva, exercício, desenvolvimento de novos hábitos); estratégias de tratamento; estratégias de prevenção; estratégias de diagnóstico. Veja tratamento.

Linha de base: dados coletados no início de uma pesquisa para todos os participantes. Devem

ser coletados dados demográficos, como idade e gênero, além das medidas específicas que serão estudadas na pesquisa e serão a base para medir mudanças nas variáveis de interesse.

Não aleatório: qualquer método que não esteja em conformidade com a definição estatística de acaso; termo usado pelos estatísticos para enfatizar a natureza de um processo fortuito ou sistemático. Veja também não casual.

Não casual: o mesmo que não aleatório.

Nível de significância: probabilidade de cometer erro tipo I, em um teste de hipóteses, com uma estatística especificada.

Número casual ou aleatório: número gerado por um processo aleatório definido.

Paciente: no contexto de pesquisa, o termo refere-se, sempre, ao paciente que participa, ou foi convidado para participar, da pesquisa.

Parâmetro: 1. em estatística, é a constante que, em uma expressão matemática, caracteriza uma população ou um processo; seu valor é, em geral, desconhecido, mas pode ser estimado. 2. em medicina clínica, é a variável cuja medida é indicativa de uma quantidade ou função que não pode ser determinada por métodos diretos. Por exemplo, a pressão sanguínea e o ritmo do pulso são parâmetros da função cardiovascular.

Participante: o mesmo que sujeito, isto é, pode ser um paciente ou apenas voluntário que participa de um estudo.

Placebo: agente farmacologicamente inativo dado a um paciente como substitutivo de um agente ativo para garantir que a resposta do paciente é explicada pela droga e não pelo fato de se supor tratado.

Poder do teste: probabilidade de rejeitar a hipótese da nulidade quando ela é falsa.

Ponto de corte: ponto, em uma sucessão ordenada de valores, que separa esses valores em duas partes.

População: todos os pacientes que poderiam, eventualmente, ser recrutados para um estudo.

Processo estocástico: diz-se do processo que depende, ou resulta, de uma variável aleatória.

Protocolo: descrição formal de todo o procedimento que será usado em um projeto específico de pesquisa.

Pseudoprocedimento: procedimento semelhante ao real, feito em alguns pacientes com a finalidade de tais pacientes (e, às vezes, o médico que faz a avaliação do ensaio) não saberem

se o procedimento adotado para esse paciente foi o real.

***p*-valor:** valor associado a uma estatística de teste que indica a probabilidade de um valor tão ou mais extremo que o observado ocorrer apenas por acaso em várias repetições de um experimento.

Randômico ou randomizado: aleatório.

Randomização: veja casualização.

Significância estatística: diz-se que houve significância estatística quando a hipótese da nulidade foi rejeitada por um teste estatístico.

***Software*:** programas de computação, manuais relacionados e demais documentos necessários para usar tais programas.

Sujeito do estudo: termo genérico que designa um indivíduo que participa de um estudo. A vantagem do termo, em relação ao termo paciente, é o fato de evitar a conotação de doença – útil nos casos em que são estudadas pessoas sadias. Veja participante.

Tamanho de amostra: 1. número de unidades experimentais do ensaio, geralmente determinado através de um cálculo, mas que também pode ser obtido por algum outro critério, como, por exemplo, estudando o que é usual na área ou recrutando as unidades disponíveis. 2. número de pacientes envolvidos em um estudo ou número de pacientes que deverão ser envolvidos em um estudo.

Tendência: (estatística) 1. diferença consistente, persistente, da estatística em relação ao parâmetro que se quer estimar. Também dita viés ou vício, traduz a palavra inglesa *bias*. 2. evolução da variável em certos sentido e direção, em geral, em função do tempo. Traduz a palavra inglesa *trend*.

Tendência: (geral) propensão, inclinação, preferência pessoal preconcebida que influencia a maneira pela qual uma medida, análise, avaliação ou um procedimento é executado ou relatado.

Teste bilateral: teste estatístico que opõe, à hipótese da nulidade, a hipótese de que existe diferença entre as populações ou os grupos em comparação.

Teste de significância: o mesmo que teste estatístico.

Teste estatístico: diz-se que foi feito um teste estatístico quando se usam dados observados e uma estatística de teste para tomar a decisão de rejeitar ou não uma hipótese e se associa a essa decisão um *p*-valor. Veja teste de significância.

Teste unilateral: teste estatístico que opõe, à hipótese da nulidade, a hipótese de que a diferença entre as populações ou os grupos em comparação está apenas à direita ou apenas à esquerda da diferença nula.

Tratamento alocado: tratamento administrado a um paciente como indicado no momento em que esse paciente decidiu participar do experimento.

Tratamento controle: droga, dispositivo ou procedimento administrado em um ensaio clínico para servir como o padrão contra o qual os tratamentos em teste são avaliados. O tratamento controle pode ser um placebo, um pseudoprocedimento, um tratamento padrão ou nenhum tratamento, dependendo do delineamento do estudo.

Tratamento padrão: maneira amplamente aceita de tratar determinada doença ou condição.

Tratamento: em estatística, regime, método ou procedimento testado em um ensaio clínico ou em um experimento, feito em qualquer área de conhecimento.

Unidade: menor unidade em que o tratamento é aplicado e cuja resposta não é afetada pelas demais unidades. Unidade básica para a coleta de dados e análises. Normalmente, refere-se a um paciente na experimentação com seres humanos, mas também pode ser material ou parte desse paciente (uma amostra de sangue, um dente) ou uma coleção de indivíduos em outros contextos (p. ex., moradores de um domicílio, uma ala de hospital). Sinônimo de unidade experimental em experimentação ou nos ensaios clínicos e de unidade observacional em estudos observacionais.

Unidade experimental: veja unidade.

Unidade observacional: veja unidade.

Variável: condição ou característica observada em cada paciente (p. ex., idade, história de infarto do miocárdio, nível de glicose no sangue) que pode assumir valores diferentes e ser observada e registrada uma ou mais vezes ao longo da pesquisa.

Variável aleatória: variável que pode assumir qualquer um de um conjunto de valores diferentes, associados, cada um, a determinada probabilidade.

Variável binária: variável que só assume um de dois valores possíveis, 0 ou 1. Veja variável dicotômica.

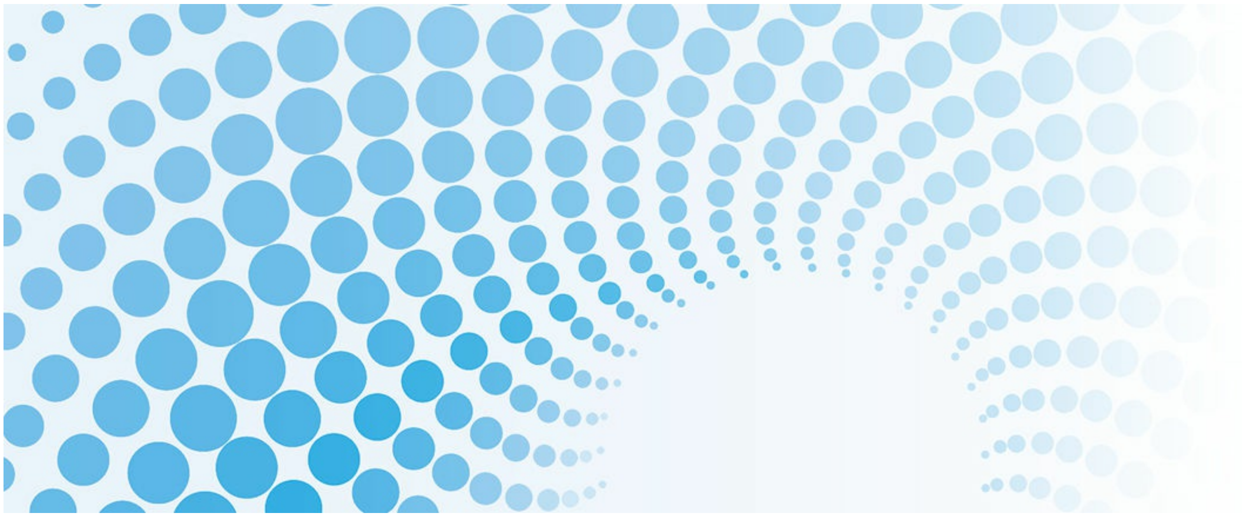
Variável contínua: variável que assume qualquer valor dentro de um intervalo especificado.

Variável dicotômica: o mesmo que variável binária.

Variável discreta: variável que só assume determinados valores em um intervalo. Veja também

variável contínua.

***Wash-out*:** suspensão temporária de medicação para remover os efeitos residuais da droga em uso pelo paciente.



A

B

C

D

E

F

G

H

I

K

L

M

N

O

P

R

S

T

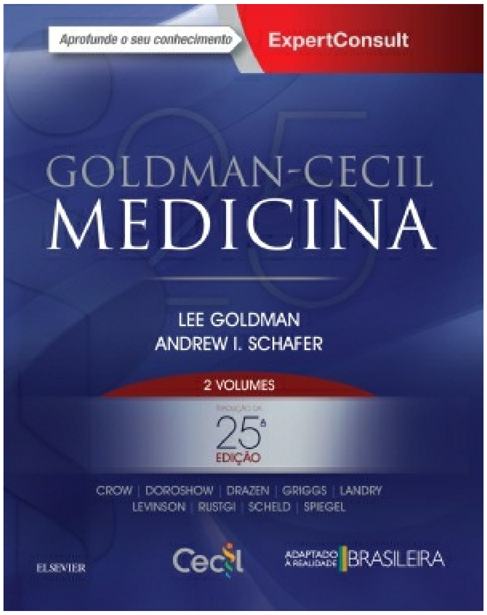
U

V

W

Z


SONIA VIEIRA
Introdução à
Bioestatística
5ª EDIÇÃO
ELSEVIER


Aprofunde o seu conhecimento
ExpertConsult
GOLDMAN-CECIL
MEDICINA
LEE GOLDMAN
ANDREW I. SCHAFER
2 VOLUMES
25ª
EDIÇÃO
CROW | DOROSHOW | DRAZEN | GRIGGS | LANDRY
LEVINSON | RUSTGI | SCHELD | SPIEGEL
ELSEVIER
Cecil
ADAPTADO À REALIDADE BRASILEIRA

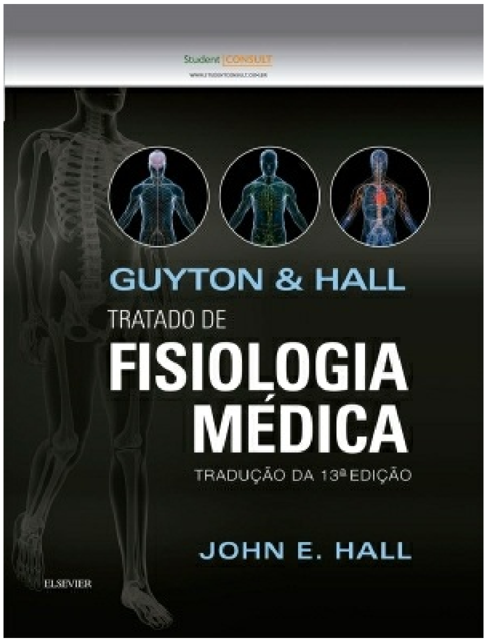
Goldman-Cecil Medicina

Goldman, Lee
9788535289947
3332 páginas

Compre agora e leia

Desde 1927, esta é a fonte mais influente do mundo sobre medicina interna. Ao comprar esta inovadora 25ª edição, além de um conteúdo ricamente adaptado à realidade brasileira sobcoordenação do Prof. Milton de Arruda Martins, você terá acesso a atualizações contínuas, disponíveis em inglês e online, que vão integrar as mais novas pesquisas e diretrizes e os tratamentos mais atuais a cada capítulo.O livro oferece orientação definitiva e imparcial sobre a avaliação e a abordagem de todas as condições clínicas da medicina moderna. As referências com base em evidência, apresentadas de maneira prática, direta e organizada, e o conteúdo interativo robusto combinam-se para tornar este recurso dinâmico a melhor e mais rápida fonte e com todas as respostas clínicas confiáveis e mais recentes de que você precisa!!• Dados epidemiológicos, estatísticas, demografia e informações atualizadas sobre a saúde no Brasil com conteúdo ricamente adaptado à realidade brasileira, contribuídos por centenas de renomados colaboradores brasileiros e coordenados pelo Dr. Milton de Arruda Martins, MD.• Texto objetivo, prático e em formato de alto padrão com recursos fáceis de usar, incluindo fluxogramas e quadros de tratamento.• Novos capítulos sobre saúde global; biologia e genética do câncer; e microbioma humano na saúde e na doença mantêm o leitor na vanguarda da medicina.• Acesso a atualizações contínuas do conteúdo, em inglês, feitas pelo editor Lee Goldman, MD– que analisa exaustivamente publicações de medicina interna e especialidades –, garantindo que o material online reflita as últimas diretrizes e traduzindo essa evidência para o tratamento.• Recursos complementares online incluem figuras, tabelas, sons cardíacos e pulmonares, tratamento e abordagem de algoritmos, referências totalmente integradas e milhares de

ilustrações e fotografias coloridas.• As mais recentes diretrizes clínicas com base em evidência atualizada que ajudam o leitor achegar a um diagnóstico definitivo e a criar o melhor plano de tratamento possível.• A abrangência do assunto com foco nos últimos desenvolvimentos em Biologia, incluindo as especificidades do diagnóstico atual, a terapia e as doses de medicação.


Student CONSULT
GUYTON & HALL
TRATADO DE
FISIOLOGIA
MÉDICA
TRADUÇÃO DA 13ª EDIÇÃO
JOHN E. HALL
ELSEVIER

Guyton E Hall Tratado De Fisiologia Médica

Hall, John E.

9788535285543
1176 páginas

A 13ª edição do Guyton & Hall Tratado de Fisiologia Médica mantém a longa tradição deste best-seller como o melhor livro-texto de Fisiologia Médica do mundo. Diferentemente de outros livros, este guia claro e de fácil compreensão tem voz autoral única e consistente e ressalta o conteúdo mais relevante para os estudantes clínicos e pré-clínicos. O texto detalhado, porém esclarecedor, é complementado por ilustrações didáticas que resumem conceitos-chave em fisiologia e fisiopatologia. • O texto com fonte maior enfatiza a informação essencial sobre como o corpo deve manter a homeostasia de modo a permanecer saudável, ao mesmo tempo em que as informações de apoio e os exemplos são detalhados com tamanho de fonte menor e destacados em lilás. • As figuras e tabelas de resumo transmitem de maneira facilitada os processos chave apresentados no texto. • Contém a nova tabela de referência rápida de valores laboratoriais padrão no final do livro. • Acréscimo do número de figuras, correlações clínicas e mecanismos moleculares e celulares importantes para a medicina clínica. • Inclui o conteúdo online em português do Student Consult, que oferece uma experiência digital aprimorada: banco de imagens, referências, perguntas e respostas e animações. Junto com a nova edição da consagrada referência mundial da fisiologia, Guyton & Hall, você também tem acesso à forma mais inovadora, simples, visual e objetiva de aprender fisiologia, o Homem Virtual, a maneira inteligente de estudar fisiologia em 3D.


TRATADO DE
GINECOLOGIA
FEBRASGO
EDITORES
César Eduardo Fernandes • Marcos Felipe Silva de Sá
COORDENADORES
Agnaldo Lopes da Silva Filho • Luciano de Melo Pompei
Rogério Bonassi Machado • Sérgio Podgaec
ELSEVIER
febrasgo

Tratado de ginecologia Febrasgo

Fernandes, César Eduardo

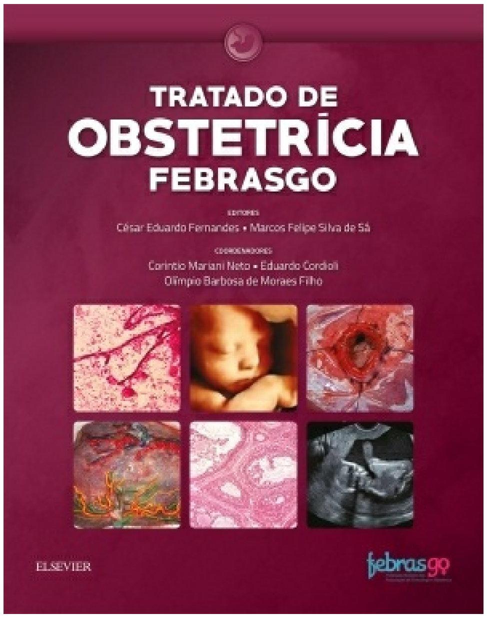
9788535292145

1024 páginas

[Compre agora e leia]

Obra referência para as provas da especialidade, certificação e recertificação na área de Ginecologia e Obstetrícia. Chancela Febrasgo. Obra referência para as provas da especialidade.

[Compre agora e leia]


TRATADO DE
OBSTETRÍCIA
FEBRASGO
EDITORES
César Eduardo Fernandes • Marcos Felipe Silva de Sá
COORDENADORES
Corintio Mariani Neto • Eduardo Cordioli
Olímpio Barbosa de Moraes Filho
ELSEVIER
febrasgo

Tratado de obstetrícia

Febrasgo

9788535292213

1024 páginas